The Phonetics of Lexical Self-Repair in Colombian Spanish Unscripted Speech

Kelly Johanna Vera Diettes

Submitted in accordance with the requirements for the degree of
PhD Linguistics and Phonetics

The University of Leeds
School of Languages, Cultures and Societies

May 2025

To Allegra,

for walking every step of this journey with me.

Declaration

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly-authored publications has been included.

The contribution of the candidate and the other authors to this work has been explicitly indicated below:

Parts of this thesis have been published in the following jointly authored work:

- Vera Diettes, K. J. and Plug, L. (2023) Temporal organization in lexical self-repair in Colombian Spanish unscripted speech. In: *Proceedings of the 20th International Congress of Phonetic Sciences*. 20th International Congress of Phonetic Sciences, 07-11 Aug 2023, Prague, Czech Republic. International Phonetic Association, Prague, Czech Republic. ISBN 978-80-908114-2-3

Candidate's contribution: Methodology, Investigation, Data collection and curation, Formal analysis, Writing – original draft, Writing – review & editing.

Co-author's contribution: Methodology, Supervision, Writing – review & editing.

The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

I acknowledge the use of ChatGPT (OpenAI, https://chat.openai.com) with code development and proofreading of sections of the final draft.

Acknowledgements

This thesis would not have been possible without the generous support of many people, to whom I wish to express my deepest gratitude.

To Leendert Plug and Gisela Tomé Lourido, my thesis supervisors. I am profoundly thankful not only for their rigorous knowledge and academic guidance but also for their humanity and unconditional support. Their patience, clarity, and commitment have been essential at every stage of this process. I always felt supported, heard, and encouraged to think more deeply. I am grateful to Leendert for being my mentor in the speech sciences; any clarity and passion I may bring to teaching phonetics and phonology will be thanks to him. Also, to Gise for allowing me to take part in her amazing Accent-Bias Project as Research Assistant. I was incredibly lucky to have you both as my supervisory team.

To the University of Leeds, for welcoming me, for giving me the opportunity to grow professionally and academically, and for allowing me to teach and conduct research there. Being part of this institution took my experience as an educator to a new level.

To the Universidad Nacional de Colombia, for granting me the study leave that made it possible to focus on this project. In particular, to the Department of Linguistics, which is my academic home, for facilitating the creation of the Unit for the Study of Interactional Speech. Without this space, the project would not have reached the same scope.

To the Ministry of Science, Technology and Innovation of Colombia – MINCIENCIAS, for funding this research through its programme for high-level training abroad. This support was fundamental in allowing me to dedicate myself fully to research and to pursue doctorate studies in such a rich academic environment.

To the participants in the project, for their willingness to take part in the recordings and for allowing their voices to become valuable data for analysis. Their collaboration and naturalness were essential in achieving the goals of this thesis.

To Allegra, for her unconditional presence at every moment of this long and demanding research journey. I thank her for sharing with me the experience of living in Leeds and for bravely adapting to a new life. Her company, energy, and joy were my greatest strengths along the way. Her presence gave meaning to every effort and always reminded me why it was all worth it.

To Felipe, for his constant love, companionship, generous knowledge, and patience in listening to me. For always being willing to discuss key topics in the thesis, engaging attentively and becoming a lucid and committed interlocutor.

To Andrea, my discourse analyst, who is also my sister. I thank her for her time, her expert insight, and above all, for her closeness. Her contributions sharpened my analyses, and her presence, at this stage, as in all others of my life, made this learning even more meaningful.

To Patricia, for her unconditional friendship and for always being there, even from the distance. Thank you for reading every word of this project, from the earliest drafts of the research proposal, over six years ago, to the final version of this thesis.

To Adriana, for accompanying and advising me during these years. Her words and guidance were a fundamental support every time I needed it.

And finally, to my family: to my father, Fabio, and my mother, Elsa, for their unwavering support; everything I am is thanks to them. This achievement belongs as much to my parents as it does to me. To my sisters, Andrea and Melissa, and to Emiliano, my nephew, for their presence,

messages, photos, and for always making me feel close. I hope I have given you a reason to be proud.

Agradecimientos

Esta tesis no habría sido posible sin el acompañamiento generoso de muchas personas, a quienes deseo expresar mi más sincero agradecimiento.

A Leendert Plug y Gisela Tomé Lourido, mis directores de tesis, les agradezco profundamente no solo por su conocimiento riguroso y guía académica, sino también por su calidad humana y su apoyo incondicional. Su paciencia, claridad y compromiso han sido fundamentales en cada etapa de este proceso. Siempre me sentí acompañada, escuchada y motivada a pensar con mayor profundidad. Le agradezco a Leendert por ser mi mentor en las ciencias del habla; todo lo que pueda enseñar con claridad y pasión sobre fonética y fonología se lo deberé a él. Y a Gise, por haberme permitido participar como asistente de investigación en su increíble proyecto sobre Sesgo de Acento. Tuve una suerte inmensa al contar con ambos como mi equipo de dirección.

A la Universidad de Leeds por acogerme, por brindarme la oportunidad de crecer profesional y académicamente, y por permitirme enseñar e investigar allí. Formar parte de esta institución llevo mi experiencia como educadora a un nuevo nivel.

A la Universidad Nacional de Colombia, por otorgarme la comisión de estudios que hizo posible que me enfocara en este proyecto. En particular, al Departamento de Lingüística, que es mi casa académica, por facilitar la creación de la Unidad de Habla Interaccional, sin este espacio el proyecto no habría tenido el mismo alcance.

Al Ministerio de Ciencia, Tecnología e Innovación de Colombia – MINCIENCIAS, por financiar esta investigación a través de su programa de formación de alto nivel en el exterior.

Este apoyo fue fundamental para poder dedicarme de lleno a la investigación y llevar a cabo estudios doctorales en un entorno académico tan rico.

A los participantes del proyecto, por su disposición conversar en las grabaciones y permitir que sus voces se convirtieran en datos valiosos para el análisis. Su colaboración y naturalidad fueron esenciales para lograr los objetivos de esta tesis.

A Allegra, por su presencia incondicional en cada momento de esta larga y demandante investigación. Le agradezco haber compartido conmigo la experiencia de vivir en Leeds, adaptándose con valentía a una nueva vida. Su compañía, energía y alegría fueron mis mayores fortalezas. Su presencia dio sentido a cada esfuerzo y me recordó, siempre, por qué vale la pena continuar.

A Felipe, por su amor y compañía constante, por su saber generoso y por escucharme con paciencia. Por estar siempre dispuesto a conversar sobre los temas clave de la tesis, involucrándose con atención y convirtiéndose en un interlocutor lúcido y comprometido.

A Andrea, mi analista del discurso, quien también es mi hermana. Le agradezco por su tiempo, por su mirada experta y, sobre todo, por su cercanía. Sus aportes hicieron más precisos mis análisis, y su compañía en esta etapa, como en todas las demás de mi vida, hizo más significativo este aprendizaje.

A Patricia, por su amistad incondicional y por estar siempre presente, incluso en la distancia. Gracias por leer cada palabra de este proyecto, desde los primeros borradores de la propuesta de investigación, hace más de seis años, hasta la última versión de esta tesis.

A Adriana, por acompañarme y aconsejarme durante estos años. Sus palabras y orientación fueron un apoyo fundamental siempre que lo necesité.

Y, por último, a mi familia. A mi papá Fabio y a mi mamá Elsa, por su apoyo desde siempre; todo lo que soy es gracias a ellos y este logro es tanto suyo como mío. A mis hermanas, Andrea y Melissa, y a Emiliano, mi sobrino, por su presencia, por los mensajes, por las fotos, por hacerme sentir siempre cerca. Espero haberles dado un motivo más de orgullo.

Abstract

This thesis investigates temporal organization and prosodic marking in unscripted lexical self-repair. It focuses on repair in Colombian Spanish, aiming to assess the extent to which the findings on the phonetics of lexical self-repair in other languages generalize to Spanish. Specifically, it aims to determine whether the distinction between trouble sources detected early in speech, in contrast to those identified late, is phonetically reflected in the studied subset of repairs; to explore the extent to which linguistic and factual errors, as opposed to appropriateness infelicities, result in phonetically distinct repairs; and to describe how sequences with and without editing terms influence repair strategies.

The thesis is divided into three parts. The first part describes the design and collection of the *Unscripted Colombian Spanish Interaction (UCSI) Corpus*, which comprises high-quality spontaneous speech recordings. The second part quantitatively examines the temporal organization of repair focusing on the time it takes to interrupt speech once a trouble source is identified (i.e., Target-to-cut-off); and the time it takes to repair once the flow of speech has bee586n interrupted (i.e., Cut-off-to-repair). Results showed that clustering the timing intervals into '*early*' and '*late*' ranges is informative, with linguistic and factual errors differing from appropriateness infelicities in the Target-to-cut-off, particularly when editing terms are present. The third part investigates prosodic marking by quantifying three acoustic parameters associated with prominence: Periodic Energy, Pitch and Speech Rate. Findings showed that while there is a preference for marking repairs in Colombian Spanish, only Periodic Energy distinguishes errors from infelicities with early interruptions unexpectedly found unmarked.

The thesis contributes to our knowledge of cross-linguistic patterns in self-repair by providing insights into the phonetic and prosodic variability within repair strategies in Colombian Spanish. The thesis also highlights the relevance of observing repair in spontaneous speech, as it uncovers further details on the interaction between language structure and real-time communication.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

REPAIR is the term that has been used in linguistics, psychology, sociology and other disciplines to cover for all the range of practices that are used to deal with problems during speech interactions (Couper-Kuhlen & Selting, 2017; Plug, 2011, 2016; Schegloff, 2007; Schegloff et al., 1977). Repair is part of our everyday exchanges since communication mishaps are as frequent as conversation itself. These disruptions, ranging in nature and severity, can trigger a variety of speaker responses aimed at resolving the trouble. Accordingly, speakers exhibit diverse strategies for identifying and addressing problematic elements in their speech. Understanding how repair practices are organized and how languages and language structure repair is of interest since by learning about them, we can better comprehend processes of language planning, monitoring and communicative recovery.

This PhD project aims at investigating the repair system described by Schegloff et al. (1977) with a focus on its phonetic realization in Colombian Spanish. Particularly, it looks at aspects of phonetic implementation such as temporal organization, speech tempo and prosodic marking within repair, to assess how these phonetic dimensions manifest in a language that remains underexplored in this domain.

Chapter 1 presents the theoretical and empirical foundations of repair and introduces how researchers have approached repair in recent years. Chapter 2 introduces *the Unscripted Colombian Spanish Interaction Corpus* (UCSI Corpus), a collection of spontaneous conversations compiled specifically for this research, which provides the empirical foundation for the analyses. Chapter 3 and Chapter 4 present experimental studies on the temporal and prosodic organization of repair in Colombian Spanish, each including its own detailed literature review, methodology, and results. Chapter 5 integrates the main findings across both studies in a general discussion, and lastly, Chapter 6 presents the thesis' final remarks.

This introductory chapter is organized as follows: Section 1.1 introduces the concept of repair and the mechanisms of repair initiation; Section 1.2 outlines the principal types of repair and elaborates on key distinctions within the system; Section 1.3 reviews the relevant literature that informs on aspects of the phonetic implementation of repair that are key to this thesis; Section 1.4 introduces the overarching research question, sub-questions and the thesis' theoretical contributions to the field; finally, Section 1.5 offers a summary of the chapter.

1.1 Repair: Initiating, Interrupting and Fixing Trouble

Schegloff describes repair as a series of 'practices for dealing with problems or troubles in speaking, hearing, and understanding the talk in conversation' (Schegloff, 2000, p. 206). These practises respond to what Schegloff describes as a *repair system*, a structured and orderly mechanism that allows speakers to manage trouble during talk-in-interaction. Functioning as a kind of 'self-righting mechanism,' the repair system enables communication to proceed despite disruptions (Schegloff, 1992, p. 1337). In the following paragraphs, I outline how this system is triggered and organized, and describe its core components.

For a repair to occur, a trouble source, such an error or infelicity, must first arise. While such disruptions are common in conversation, they are not always addressed. They can go unnoticed, or participants in the interaction can just simply let them go untreated. Since this thesis focuses specifically on cases where speakers *do* initiate repair, unrepaired trouble sources will not be discussed further. The analyses presented in this dissertation centre instead on instances where speakers judge their problematic productions to be worth correcting.

Throughout this thesis, I present examples of repair sequences extracted from the UCSI Corpus. These examples are used to illustrate key aspects of repair operations, allowing for authentic, data-grounded discussion rather than relying on constructed examples. All examples are transcribed using conversation-analytic conventions, following Hepburn and Bolden (2017). In the transcripts, the problematic item (i.e., the target of correction) and the repair (i.e., the correction itself) are shown in **bold**. In complex repair sequences, the turn containing the repair initiation is marked with an arrow (→); and (–) denotes the point of cut-off. Note that when required, examples and descriptions will include broad phonetic transcriptions enclosed in squared brackets.

In (1) Speaker A notices a problem in their own speech. There, the issue comes with the choice between the prepositions *de* 'of' and *con* 'with'. The speaker initially produces the utterance using *of* as first choice, to end up interrupting the sequence at the next word *influx*, which we can see only partially pronounced as *aflue-* [ˈafl̯ue]. After interrupting, the correction comes in by changing the preposition from *of* to *with,* which the speaker decides to be a better fit for the noun. Additionally, the speaker adds the adjective *great* before the noun to add emphasis, leaving us with the adjusted utterance *with great influx*. From this example we can name several important components of the repair sequence. Firstly, the preposition *of* is the REPARANDUM, which is how I will refer to the error or infelicity throughout the thesis (with REPARANDA used as the plural form). Secondly, *with great* is the outcome of the repair operation, therefore, the REPAIR itself (i.e., solution or correction) (Schegloff, 1992, 2000). Also, we can note that the cut-off; that is, the moment of interruption, happened after the reparandum was fully produced.

Therefore, *de* 'of' in (1) is an instance of a COMPLETE reparandum. This type of reparanda contrasts with cases in which the cut-off happens within the reparandum itself, which leave us with INCOMPLETE reparanda items, as seen in example (2). Finally, the reparandum is identified and acted upon, within the same turn, by A, who is the same speaker who produced the problematic piece of speech. In consequence, (1) and (2) are instances of SELF-INITIATED REPAIR.

(1) SELF-INITIATED REPAIR with a COMPLETE reparandum

Spanish (Vera Diettes, UCSI Corpus, 9, sp_003_004_conv_kvpart5) [1]

01	A:	**de** aflue- **con gran** afluencia ***of*** *influ-* ***with great*** *influx*

(2) SELF-INITIATED REPAIR with an INCOMPLETE reparandum and EDITING TERMS

Spanish (Vera Diettes, UCSI Corpus 382, sp_021_022_dpix1_kv_part1)

01	A:	y un **vestid- pues sí,** como un **traje** naranja con falda *and a* ***dress****, yes, like a* ***gown****, orange with a skirt*

(2) is another example of a self-initiated repair. While describing an outfit, Speaker A first talks about a 'dress' to, then, change the piece of clothing to a 'gown'. The Spanish word for

[1] All examples are drawn from the Unscripted Colombian Spanish Interaction (UCSI) Corpus, compiled and annotated by the author. Each example is identified using the standardized naming convention described in Section 2.6.1, which includes project acronym, participant codes, task type, fragment number, and researcher initials.

dress, *vestido* /besˈtido/ is partially articulated as [besˈtið̞], resulting in an INCOMPLETE reparandum. Another important part of this repair sequence is the presence of editing expressions in between the reparandum and the repair. According to Levelt (1983, p. 41), when trouble is identified, it is followed by hesitation, pausing and, possibly, the use of EDITING TERMS, as I will refer to them in this thesis, following Levelt's (1983) description. Editing terms are expressions such as 'uh', 'sorry' or 'I mean'. In example (2) we can read the sequence of words *pues sí, como* which could be interpreted in English as 'yes, like'[2]. Levelt (1983), later (Nooteboom, 2010; Nooteboom, 2005b) and Plug (2016), looked at editing terms in repair and identified them as functional and important part of repair sequences.

Example (3) illustrates OTHER-INITIATED REPAIR. There, the interlocutor (B) initiates the repair by questioning the description provided by Speaker A regarding two colours in an object and even offers an alternative as a possible correction. In their next turn, A accepts the repair by changing the colour initially described from *white* to *grey*. B subsequently confirms this choice as correct.

As seen above, we can have repair sequences that are initiated by the speakers themselves, or by the others. To complete the picture, we also need to observe who is the person that ends up repairing after a reparandum is first identified. Again, the classification between *self* and *other* helps us distinguish the identity of the repairer; examples (4) and (5), below, illustrate how this distinction works.

---

[2] As noted by Levelt (1983), editing terms do not correspond straightforwardly across languages. For this reason, the Spanish-to-English renditions of editing terms presented here, and elsewhere in the thesis, are not direct translations, but rather interpretations of their pragmatic meaning.

(3) OTHER-INITIATED REPAIR

Spanish (Vera Diettes, UCSI Corpus, sp_009_010_dpix_1_kvpart3)

01	A:	es roja con **blanco**
		*it is red and **white***
02	B: →	¿blanco como el granero? o ¿más bien **gris**?
		*white as the barn? Or is it more like **grey**?*
03	A:	**gris**
		grey
04	B:	sí
		yes

In (4) the two speakers are talking about what they see on a street. A is describing what they see on the zebra crossing but mistakenly evokes a *giraffe* instead of a *zebra,* when talking about the crossing. The error is immediately recognized, and A interrupts the articulation of the word, resulting in an incomplete reparandum. A first identifies the problem and tries to find the correct word, as we can see in line 1, but fails. Immediately afterwards, B comes up with the repair *zebra.* Both speakers laugh at the exchange before A rephrases the whole sentence and produces a new utterance with the right word. It is important to note that although A ultimately repeats the sentence with the appropriate term, this repetition does not constitute the repair itself, since it was Speaker B who resolved the trouble source, giving us an example of an OTHER-REPAIR. More specifically, because the repair was initiated by A but completed by B, (4) is an instance of a SELF-INITIATED, OTHER-REPAIR.

(4) SELF-INITIATED, OTHER-REPAIR

Spanish (Vera Diettes, UCSI Corpus, 9, sp_009_010_dpix_2_kvpart1)

01	A:	Sobre la **jira-;** la qué, la ((laughing)) *On the* ***gira-*** *(crossing) ((laughing)); the what? The...*
02	B: →	La **cebra** *The* ***zebra (crossing)***
03	A:	((laughing)) la **cebra,** dije la jirafa. *((laughing)) The* ***zebra (crossing),*** *I said giraffe.*
04	B:	((laughing))
05	A:	Sobre la cebra, hay un animal. ((laughing)) *On the zebra crossing, there is an animal. ((laughing))*

Example (5) is also an instance of other-repair, but oppositely to (4), in (5) the speaker does not recognize the problem until after B rephrases the entire utterance originally produced by A. In doing so, B replaces *representative* (the reparandum) with *representee* (the repair), which is the correct term. The exchange finishes in line 3, with speaker A repeating the utterance, but this time correcting the error, and even adding an excuse word. Altogether, (5) is an instance of an OTHER-INITIATED, OTHER-REPAIR.

(5) OTHER-INITIATED, OTHER-REPAIR

Spanish (Vera Diettes, UCSI Corpus, sp_009_010_conv_kvpart34)

01	A:	Su **representante** no es inocente. *Your* ***representative*** *is not innocent.*
02	B: →	Su **representado** no es inocente. *Your* ***representee*** *is not innocent.*
03	A:	Su representado, perdón, no es inocente. *Your representee, sorry, is not innocent*

Returning to self-initiated repair, as seen previously, speakers may both identify the problem and produce the correction themselves. These are instances of SELF-INITIATED, SELF-REPAIR, and they represent the most common type of repair observed in everyday conversation. Example (6), as well as (1) and (2) presented earlier, are all instances of self-initiated, self-repair.

(6) SELF-INITIATED, SELF-REPAIR

Spanish (Vera Diettes, UCSI Corpus, 9, sp_025_026_dpix3_kv_part4)

01	A:	Y unas **sandalias,** unas **chanclas.** *And some* ***sandals,*** *some* ***flip-flops.***

To complete the overview on who triggers the repair and who fixes it, the final combination is OTHER-INITIATED, SELF-REPAIR. These are cases in which the interlocutor brings the attention to the trouble source, without offering solutions to the issue. Instead, the speaker who produced the reparandum also provides the repair, prompted by the interlocutor's intervention. (7) is an illustration of this type of exchanges. There, Speaker A uses the name of a famous dog to refer to an actual dog that was sitting on a chair in a picture. Speaker B does not understand the reference and asks about *what* was on the chair. Next, Speaker A rephrases the sentence clarifying it was a dog, like the famous one, what was on the chair.

(7) OTHER-INITIATED, SELF-REPAIR

Spanish (Vera Diettes, UCSI Corpus, 9, sp_025_026_dpix3_kv_part1)

01	A:	Y hay **un 'Chems'** ahí en la silla.
		*There is a **'Chems'** there in the chair.*
02	B:	Solo dice, solo dice ahora.
	→	¿En la silla qué hay?
		It just says, it just says now.
		What is there on the chair?
03	A:	**Un perrito** así como 'Chems'.
		A little dog like 'Chems'.

To sum up, the combination between the two options among who starts repairs and who resolves the trouble, give rise to the main four repair types that we can distinguish (Couper-Kuhlen & Selting, 2017; Schegloff et al., 1977):

1. SELF-INITIATED, SELF-REPAIR; in which the same speaker that identifies the trouble produces the repair, as illustrated in (6), and also the case for (1) and (2);

2. OTHER-INITIATED, SELF-REPAIR, where the speaker produces the repair after the trouble is flagged by the interlocutor, as in (7);
3. SELF-INITIATED, OTHER-REPAIR, which occurs when the interlocutor provides the repair, even though the problem was initially identified by the speaker, as in (4); and
4. OTHER-INITIATED, OTHER-REPAIR, here the interlocutor both identifies the trouble source and produces the repair, as exemplified in (5).

Research on repair has approached all these configurations in an effort to address the different questions that have come up during years of research on the topic. For the time being, the way in which repair is organized seems to hold in everyday talk-in- and interactional contexts. As such, this typology of four repair types is often described as 'quasi-universal' (Couper-Kuhlen & Selting, 2017, p. 116).

1.2 The Emergence of Different Repair Subtypes

So far, we have seen that repairs can be categorized based on who initiates and who completes them. In addition to this interactional dimension, repairs can also be classified according to the nature of the trouble they address. If we inspect their structure from a linguistic perspective, they can be phonological, lexical, or grammatical. Furthermore, from a pragmatic view, repairs can be the response to infelicities that go against social appropriateness (Nooteboom, 2005b, p. 44), or just simply factual errors.

In this section, I introduce the most common repair subtypes, with a focus on their phonetic forms and communicative functions. To that end, I will first present phonological and lexical repairs, to later discuss how these and further classifications have proven useful for understanding the strategies speakers use to resolve trouble, as well as the role of repair in speech self-monitoring.

1.2.1 Phonological and Lexical Repairs

The distinction between phonological and lexical repairs is commonly found in the literature as an informative subdivision for the study of repair (i.e., Levelt, 1983; Levelt & Cutler, 1983; Levelt et al., 1999; Nooteboom, 2010; Nooteboom, 2005a, 2005b; Plug, 2011, 2015a, 2016). PHONOLOGICAL REPAIRS are those errors in which 'phonemes are misplaced' (Nooteboom, 2005b, p. 168). Following Postma (2000) and Plug and Carter (2014), throughout this thesis, I will refer to them as phonological repairs, but they can also be found in literature as *sound-form* error repairs (Levelt & Cutler, 1983); *phonetic* error repairs (Bredart, 1991) or *speech* error repairs (Nooteboom, 2010; Nooteboom & Quené, 2019). Example (8), below, is an illustration of a phonological self-initiated, self-repair, where a speaker mispronounces the Spanish word *carro* 'car'. The reparandum is an incomplete erroneous realization of the word ([ke-]), and the repair is its correct pronunciation. Generally, phonological repairs involve the same lexical item for both the reparandum and the repair.

(8) PHONOLOGICAL, self-initiated, self-repair

Spanish (Vera Diettes sp_017_018_dpix1_kv_part1)

01	A:	Un **ca-** [k**e**-]; Un **carro** [ˈk**a**.ro]
		A ***ca-****; A* ***car****.*

LEXICAL REPAIRS, on the other hand, are instances in which a speaker rejects one lexical choice in favour of another (Plug, 2015a, p. 80). These reformulations may stem from the need for greater accuracy, clarity, or contextual appropriateness, points I will return to in more detail later. (9) is an example of a lexical self-initiated, self-repair, where the speaker selected a wrong lexical choice, interrupted the flow of speech and made a repair by restarting the phrase and replacing the error with the right word. It is also worth noting that, as opposed to example (8) above, in (9) the reparandum is *complete*, and the moment of the cut-off comes afterward, in the word immediately following the reparandum. This highlights that the point of cut-off can vary from one repair to another and does not always occur within, or immediately after, the reparandum. In some cases, the flow of speech continues briefly before the speaker initiates the interruption, suggesting a short delay between recognition of trouble and response (Blackmer & Mitton, 1991; Seyfeddinipur et al., 2008).

(9) LEXICAL, self-initiated, self-repair

Spanish (Vera Diettes, UCSI Corpus, 65, sp_007_008_dpix_1_kv_part4)

01	A:	tiene una **bolsa** r- una **camisa** rosada como con un rayo rojo *He has a **bag** p-, a pink **shirt,** with like a red lighting*

1.2.2 Issues of Pragmatic Felicity and Inaccuracies in Lexical Repair

We have seen that a mistaken target word can sometimes be simply misplaced. However, in other cases the problem obeys to a difficulty organizing more complex structures or fitting them appropriately to the context (Searle, 1969). We as speakers not only aim to produce utterances that are linguistically accurate, grammatically and lexically correct, but also, to ensure that the information is factually true and pragmatically appropriate for the social context in which it is being used. When we notice that any of these dimensions goes wrong, we are likely to attempt a repair as soon as the problem is recognized.

More recent research on repair follows Levelt (1983) and Levelt and Cutler (1983) in distinguishing issues of pragmatic felicity from linguistic and factual accuracy. When a repair addresses a trouble source that is linguistically or factually inaccurate, it is classified as an ERROR REPAIR, whereas repairs prompted by pragmatically infelicitous constructions are referred to as APPROPRIATENESS REPAIRS (Levelt, 1983, pp. 51-55).

The excerpts below further illustrate the distinction between appropriateness and error repairs. First, example (10) shows a lexical self-initiated, self-repair repair, which can be classified as an error repair. There the speaker mismatches a number in the description given. Also, in (10), note that the editing term *bueno* 'well' has surfaced in between the reparandum and the repair.

Example (11) illustrates an appropriateness repair, a lexical, self-initiated, self-repair, where the speaker initially uses a borrowed English word to refer to a performance or play. After a pause, the speaker replaces it with the equivalent term in Spanish, completing the utterance entirely in the local language. In Colombian Spanish, English borrowings are widely accepted; therefore, the use of the English word in this case does not constitute an actual error. It is likely that the speaker reconsidered whether this was the best way to express the idea, and ultimately opted for the Spanish equivalent, which may have been perceived as more contextually appropriate. What is particularly noteworthy in this example is the delay between the initial lexical choice and the production of the repair. Although the reparandum was fully articulated, the cut-off occurred several words later, well after the trouble source itself. This is relevant to the present thesis, as it shows that the monitoring processes involved in self-repair can vary considerably. In this case, the speaker appears to have taken significant time to assess whether the lexical item originally selected was the most suitable for the context.

(10) ERROR, lexical, self-initiated, self-repair

Spanish (Vera Diettes, UCSI Corpus, 54, sp_009_010_dpix_1_kvpart1)

01	B:	¿Cuántos niños hay cerca a la canasta? *How many children are there near the basket?*
02	A: →	**cuatro**, bueno, **tres**; hay un niño que está cerca pero está mirando hacia otro lado ***four**, well, **three**; there is one kid that is near but is looking to the other side*

(11) APPROPRIATENESS, lexical, self-initiated, self-repair

Spanish (Vera Diettes, UCSI Corpus, 149, sp_021_022_ conv_kv_part4)

01	A:	Ese **performance** está, como de alguna forma, tan bien montado-; esa [ˈe.sa:] **obra de teatro** está tan bien montada. *That **performance** is, in a way, so well put up-; that **performance** is so well put up.*

Although Levelt (1983) and later Levelt and Cutler (1983) acknowledged the distinction between linguistic and factual errors, their primary focus was on the broader classification between appropriateness and error repairs. Building on this foundation, Plug and Carter (2013) and Plug (2016) examined the sub-classification of errors in greater detail, proposing that the distinction between linguistic and factual errors may have implications for the phonetic realization of both the reparandum and the repair.

According to Plug, LINGUISTIC ERROR REPAIRS are those in which there is a construction that is considered grammatically inappropriate, while in FACTUAL ERROR REPAIRS speakers' first choice of a semantic target is incorrect (Plug, 2016, p. 519). Example (12) illustrates a lexical self-initiated, self-repair of a linguistic error. There, the speaker constructs an ungrammatical question by using a noun in the verb position within the verb phrase; and (13) presents a lexical self-initiated, self-repair of a factual error in which the speaker revises an animal classification, correcting *savage* to *wild.*

(12) LINGUISTIC ERROR, UCSI Corpus, 146, lexical, self-initiated, self-repair

Spanish (Vera Diettes sp_023_024_conv_kv_part26)

01	A:	En qué **agravante**-; ¿en qué le **empeora** eso la vida a la persona? *How does this **aggravation**-; how does this **worsen** someone's life?*

(13) FACTUAL ERROR, lexical, self-initiated, self-repair

Spanish (Vera Diettes, UCSI Corpus, 9, sp_019_020_conv_kv_part9)

01	A:	Los animales **salvas**-; **silvestres**, ¿no? ***Savag-** (animals); **wild** animals, right?*

These distinctions among repair types and subtypes lay the groundwork for examining how repair is realized phonetically. The following section introduces timing and temporal organization, two key dimensions through which we can observe how speakers manage trouble in real time.

1.3 Phonetic Features of Repair

1.3.1 Timing and Temporal Organization

When observing how speech is organized in repair, one of the relevant phonetic features that has been studied is timing. Studying timing through the measurement of duration intervals extracted from the components of the reparandum-repair sequence can shed light on how speakers deal with speech problems as they arise. For instance, timing can reveal whether it takes longer to resolve a linguistic error than a factual one, or whether correcting an appropriateness infelicity is more temporally demanding than addressing a clear-cut error. In sum, duration measurements offer insight into the temporal organization of repair, and can help determine whether different types and subtypes of repair are associated with distinct timing patterns.

Within this view, several hypotheses have emerged from the study of repair timing as captured through duration measurements. I will outline a few of them briefly here; Chapter 3 provides a more detailed review of the relevant literature.

Nooteboom (2010) proposed that phonological self-repairs may be monitored through distinct cognitive processes, depending on whether the trouble is detected before articulation (inner speech) or after articulation (overt speech). He argues that monitoring for speech errors in

INNER SPEECH[3] (pre-articulatory) is under some sort of time pressure due to the need to avoid errors from being articulated and, therefore, becoming public. Quite the opposite, monitoring for speech errors in OVERT SPEECH (post-articulatory) is not under such pressure, basically because the problem at this point has already become evident. Under this perspective, the duration between the reparandum offset and the onset of the repair could shorten in response to the proposed idea of 'time pressure to prevent embarrassment' (Nooteboom, 2010, p. 228). It is important to note, however, that this inner–overt distinction is not directly observable; rather, it is an interpretive framework grounded in the temporal characteristics of repair sequences. Based on these patterns, Nooteboom infers that certain timing profiles (e.g., rapid cut-offs followed by quick repairs) are likely indicative of early detection, whereas others (e.g., late interruptions with longer delays) may indicate post-articulatory monitoring.

Plug and Carter (2014), in their study of spontaneous phonological repairs, found patterns broadly consistent with Nooteboom's (2010) analysis: repairs that are initiated early tend to be completed quickly, while later-initiated repairs are typically slower to resolve.

Plug's (2016) focuses on the relationship between offset timing, repair tempo, and several predictors including semantic subtype, lexical frequency, and reparandum completeness. While his findings offer only limited support for Nooteboom's (2010) inner vs. overt monitoring

[3] I will use the term 'Inner Speech' throughout the dissertation following Levelt (1989) and Levelt et al.'s (1999) definition of this early stage of error detection in speech monitoring. However, the reader should keep in mind that phonological self-repairs may be monitored through distinct cognitive processes, depending on whether the trouble is detected before articulation (inner speech) or after articulation (overt speech) more recent accounts have suggested that 'inner speech' differs from 'internal speech' in that the latter corresponds to a stage of speech preparation, while the former is to remain silent. Thus 'inner speech' does not exhibit certain low level phonetic features that 'internal speech' does. (c.f. Oppenheim, G. M., & Dell, G. S., 2008)

hypothesis, they suggest that timing characteristics are partially influenced by the point at which the reparandum is interrupted, with earlier offsets tending to yield faster repairs. Plug also explored whether dividing error repairs into factual and linguistic subtypes, alongside the broader appropriateness vs. error distinction, could account for differences in temporal organization (Plug, 2016, pp. 518-519). However, the effects of this semantic classification were relatively weak and inconsistent across measures. His analysis ultimately highlights that reparandum completeness and lexical frequency offer more robust explanatory power than semantics alone, and calls for future research to integrate multiple dimensions of repair when modelling self-monitoring behaviour.

Taken together, these findings underscore the need for further research with larger and more balanced datasets, capable of capturing the range of repair types and subtypes. Such studies would allow for more refined modelling of temporal organization in repair (Plug, 2016, p. 537).

1.3.2 Prosodic Marking

In addition to timing, another dimension of interest in the phonetic study of repair is prosodic marking, a phenomenon that includes temporal cues such as duration and tempo, alongside pitch and intensity, to enhance the acoustic salience of the repair. It might be expected that speakers, upon detecting an error or infelicity, would emphasize the repair to contrast it with the original, thereby guiding the listener toward the correct interpretation. However, findings from previous research challenge the assumption that repairs are always prosodically marked. I will briefly account for results on prosodic marking in repair in the next paragraphs; however, a thorough discussion on the state of the art in this matter will be offered in Chapter 4.

Both Cutler (1983) and Levelt and Cutler (1983) explored the distinction between error and appropriateness in lexical self-repairs by observing prosodic marking. They found evidence to support the idea that speakers' preference for marking responded to the need to contrast information, proposing that error repairs more often required marking, while appropriateness ones needed elaboration rather than contrast (Levelt & Cutler, 1983, p. 212).

More recently, Plug and Carter (2013) investigated this distinction using acoustic data, measuring prosodic parameters such as pitch and intensity. Unlike Levelt and Cutler (1983), whose classification was based on auditory judgements, Plug and Carter found no significant effect of the error–appropriateness distinction on prosodic marking overall. However, when error repairs were further subdivided, they observed that factual error repairs were more frequently marked than either linguistic or appropriateness repairs. This suggests that a more fine-grained semantic categorization may be necessary to account for patterns in prosodic behaviour.

Nooteboom (2010) offered another perspective by linking prosodic marking to the timing of error detection. He found that phonological repairs with incomplete reparanda, those which are interrupted early, tended to be produced with higher pitch and intensity than completed ones. This supports the idea that interrupted, or early-detected repairs are initiated through inner speech monitoring, while completed, late-detected repairs are initiated through overt speech monitoring. Concretely, what this suggests for prosodic marking in repair is that when an error is identified in inner speech, it is produced as conspicuously as possible, drawing the interlocutor's attention towards the repair by increasing pitch and intensity. Conversely, if the trouble source is identified during overt speech monitoring, after the error is fully articulated and made public, speakers may not feel the need to mark the repair prosodically, as doing so is unlikely to redirect the interlocutor's attention away from the already spoken error.

1.3.3 Speech Tempo: Expansion and Reduction in Prosodic Marking

In addition to pitch and intensity, speech tempo has also been investigated as a potential cue to prosodic marking. While tempo is sometimes discussed independently of prosodic structure, this thesis follows Plug (2011, 2016) in treating articulation rate, and by extension, phonetic reduction and expansion, as part of the marking strategies available to speakers during repair. That is, the rate at which a repair is articulated may affect its perceptual salience and contribute to whether the correction is perceived as prominent or minimized.

This view aligns with the *Hyper*- and *Hypo*-articulation (H&H) theory by Lindblom (1990), which suggests that speakers adjust the speed of their speech productions based on the perceptual needs of the listener and the contextual information available. HYPO-ARTICULATED SPEECH refers to instances in which segments are produced with increased speed and reduced articulatory precision, resulting in phonetic reduction. Conversely, HYPER-ARTICULATED SPEECH involves slower, more deliberate production, often in contexts without time pressure, leading to longer segment durations and clearer articulation.

Plug (2011) examined phonetic reduction in self-initiated repairs and found a consistent pattern of hypo-articulation, where the repair was produced with increased tempo. These findings challenged the expectation that repairs introducing new information would be expanded, therefore, emphasized. Lexical frequency and the error-appropriateness distinction were tested as potential explanatory variables, but neither accounted for the reduction observed. Plug proposed that this pattern may be linked to the informational redundancy of self-repair in spontaneous conversation, or alternatively, to a face-saving strategy, where speakers minimize the perceptual impact of an error by articulating the correction somehow less prominently (Plug, 2011, p. 269).

Plug (2016) further explored tempo alongside temporal organization. Comparing the articulation rates of repairs and their corresponding reparanda, he reported that repair stretches were typically articulated at a faster rate. Reparandum completeness was included as a predictor with findings indicating that repairs following incomplete reparanda were produced more rapidly than those following complete ones (Plug, 2016, p. 533). Since completeness can be taken as a proxy for *early* versus *late* repairs (assuming that incomplete reparanda reflect early detection), these results align with Nooteboom's (2010) findings on early repairs being sorted faster than late repairs.

1.4 General and Specific Objectives

Despite decades of research on repair in conversation, much of what is known about its phonetic and interactional organization comes from studies of Germanic languages, particularly English and Dutch. In contrast, systematic, phonetic investigations of repair in non-Germanic languages remain scarce. The general objective of this thesis is to address that gap by investigating a set of phonetic features involved in the repair of lexical self-repairs in Colombian Spanish. Lexical self-repair was selected as the primary focus of research because it is both frequent in everyday conversation and analytically rich: it captures how speakers respond in real time to problems of speech planning, formulation, or pragmatic appropriateness. The analysis is based on a quantitative approach, using Bayesian statistical modelling (See 3.3.5 for details) to examine the phonetic features involved in repair, particularly those related to temporal organization and prosodic marking. While the study remains phonetic in focus, interpretation of the findings is informed by relevant insights from discourse and interactional research.

To fulfil the general objective, this thesis also aims to contribute to the field by providing a new corpus of unscripted Colombian Spanish speech, referred to throughout as the UCSI Corpus (See Chapter 2). The interactive tasks used to collect the data, described in detail in Sections 2.3.1 and 2.3.2 were chosen given their potential to elicit numerous instances of lexical self-repair, including both factual and linguistic errors. Therefore, the UCSI Corpus constitutes not only a core objective of this PhD project in its own but also forms the empirical foundation for the repair collections analysed throughout the thesis.

Although the phonetic objectives of this project focus specifically on lexical self-repair, I provide an account of the repair instances found in the UCSI Corpus (Section, 2.7.3). This broader analysis allows for an exploration of repair in Colombian Spanish in terms of who initiates and who completes the repair, the patterns observed in the emergence of phonological and lexical repair subtypes, and the relative frequency of linguistic and factual errors compared to appropriateness infelicities.

1.5 Research Questions

This dissertation investigates lexical self-repair in Colombian Spanish, focusing on the temporal and prosodic features of error repairs (linguistic and factual) and appropriateness repairs, while also considering aspects of their discourse context. The overarching aim is to assess the extent to which findings on the phonetics of lexical self-repair in other languages generalize to Colombian Spanish. The following specific research questions are addressed:

(1) To what extent does the distinction between repairs identified in *inner* and *overt* speech features in the phonetics of repair?

(2) To what extent are linguistic and factual errors and appropriateness repairs phonetically distinct?

(3) To what extent are repairs with and without editing terms phonetically distinct?

By answering the research questions outlined above, this thesis aims to contribute to the field in several ways. First, I seek to advance our understanding of the linguistic organization of repair as a system by describing how it is phonetically realized in the variety of Spanish under study. Second, by reporting the results of this thesis, I aim to contrast new empirical data with findings previously reported for other languages, to shed light on how repair may function cross-linguistically. Finally, I consider whether repair is context-sensitive, exploring the extent to which it reflects conventions that speakers establish in everyday interaction, as well as how these conventions may be shaped by language-specific and cultural norms.

1.6 Summary

To study repair in Spanish is particularly valuable given that the repair system remains under-investigated across the world's languages. As Couper-Kuhlen and Selting (2017) have noted, it is still an open question whether the patterns described in well-studied Germanic languages, particularly English and Dutch, are generalizable to other language families, structures, and cultures. This thesis addresses that gap by examining how repair operates in Colombian Spanish, a major Latin American variety. It does so by analysing the phonetic realization of repair types by focusing on their temporal and prosodic characteristics and examining how these relate to structural processes of monitoring, error detection, and self-correction as described in prior research (e.g., Levelt, 1983; Levelt & Cutler, 1983; Nooteboom, 2010; Plug, 2011, 2015a).

Given that Spanish is the second most spoken language in the world by number of native speakers with 483 million speakers (Instituto Cervantes, 2019), a detailed account of repair in this language contributes meaningfully to ongoing discussions in the typology of language use and interaction. Within the Colombian context, this research represents a pioneering step in the phonetic study of repair, and it opens the door to new questions and hypotheses that may support the development of a broader scientific inquiry into Spanish-language interaction across registers and contexts.

Chapter 2

The Unscripted Colombian Spanish Interaction Corpus: UCSI Corpus

Intrinsic to this thesis is the Unscripted Colombian Spanish Interaction (UCSI) Corpus, which was collected as part of this PhD project. This corpus provides a new, high-quality audio recording resource of unscripted Colombian Spanish speech. Beyond offering a robust empirical foundation for the present study of self-initiated repair, the UCSI Corpus also opens the door to a broad range of phonetic, pragmatic, and sociolinguistic investigations in Colombian Spanish.

Given the aims of this research, collecting the right data was of equal importance as the research objectives themselves. As Enfield et al. (2013, p. 344) noted ‘grammarians do not describe the structures of repair mainly because there is no tradition of such description in linguistics, and partly because linguists have tended not to work with the one kind of data in which these structures can be found: i.e., spontaneous talk in conversational interaction’. This thesis confronts that gap by developing a spontaneous speech corpus that is tailored to enable the analysis of repair structures in natural interaction, precisely the environment in which they arise.

Corpora that serve multiple research agendas provide long-term value to the field. By making the UCSI Corpus openly available, this project aims to foster new research not only on repair but also on broader features of Colombian Spanish. This includes studies of articulation, turn-taking, discourse markers, or variation across dialects and speakers. Furthermore, the corpus contributes to the development of linguistic research in Colombia, a country marked by rich linguistic diversity but limited access to spontaneous data resources. Moreover, Latin American varieties of Spanish, Colombian Spanish among them, remain relatively underrepresented in corpus-based research. This corpus provides a valuable resource not only for the study of self-initiated repair in spontaneous speech, but also for future investigations into the structural and interactional properties of Spanish as spoken in the region.

Lastly, Replicability as well as new opportunities for research will also be favoured by creating corpora that can be used in subsequent projects. This approach follows successful precedents. That is the case of the Ernestus corpus of spontaneous Dutch[4] (Ernestus, 2000) which has provided a framework for several projects investigating diverse linguistic features of Dutch (i.e., Plug, 2010; Schuppler et al., 2011).

In order to support the objectives of this thesis and facilitate future research, the corpus was constructed to meet the following criteria:

1) The corpus contains unscripted, spontaneous speech.
2) The corpus represents Colombian Spanish, so it contains speech from the main regional dialects spoken in Colombia.
3) It comprises a corpus rich in self-initiated repair sequences, enabling quantitative analysis across repair types and contexts[5].

[4] The Ernestus Corpus of Spontaneous Dutch (ECSD) consists of high-quality recordings of spontaneous conversations between Dutch speakers, accompanied by orthographic and phonemic transcriptions in Praat TextGrid format. Access is available to researchers by request via Radboud University: https://www.mirjamernestus.nl/Ernestus/ECSD.

[5] While the number of repair sequences per speaker varies, the corpus overall is well-suited for investigating both population-level and speaker-specific effects. The use of a Bayesian statistical framework, which is particularly robust for small or unbalanced sample sizes, further enhances the analytical power of the study. This design allows for meaningful generalizations despite the natural limitations that arise when working with spontaneous conversational data.

4) Each speaker is represented by extended stretches of speech, enabling studies of inter-speaker variation and supporting future research beyond the scope of this project.

In this chapter, I describe the key linguistic features of Colombian Spanish considered during corpus design, present the structure and contents of the UCSI Corpus, and evaluate its adequacy for studying lexical self-repair and beyond.

2.1 Colombian Spanish

Colombia is located at the extreme north of South America, at latitude 4° North and longitude 73° West (Instituto Geográfico Agustín Codazzi [IGAC], 2020). It shares boundaries with Panama (Central America) to the northwest, with Venezuela to the east, with Ecuador to the southwest and with Peru to the south. Colombia also has maritime limits in the Caribbean Sea with Nicaragua, Costa Rica, Honduras, Jamaica, Haiti, and the Dominican Republic. The country has a population of 52,695,952[6] inhabitants (Departamento Administrativo Nacional de Estadística [DANE], 2025).

Colombian Spanish and its dialects have been the focus of sustained linguistic attention over the past several decades. Most of the work started around the mid-twentieth century with an initiative to collect speech data samples for the creation of a national linguistic atlas led by Flórez (1961). Research has advanced rapidly in recent years, with investigations that have studied some of the most common phonological processes that occur in the Colombian Spanish,

[6] Based on the most recent population estimate published by DANE, updated in 2024.

and which are, to a certain extent, shared with other Latin American varieties. One example is *yeísmo*, a widespread phonological process in which the phonemes /ʎ/ and /ɟ/ merge in favour of /ɟ/, eliminating the distinction between words such as *pollo* /ˈpoʎo/ ('chicken') and *poyo* /ˈpoɟo/ ('stone bench'), both of which are realized as /ˈpoɟo/ in *yeísta* varieties (e.g., Orduz Navarrete, 2013). Another phenomenon is the aspiration of fricatives, investigated by Espejo Olaya (2016) and File-Muriel (2009), among others.

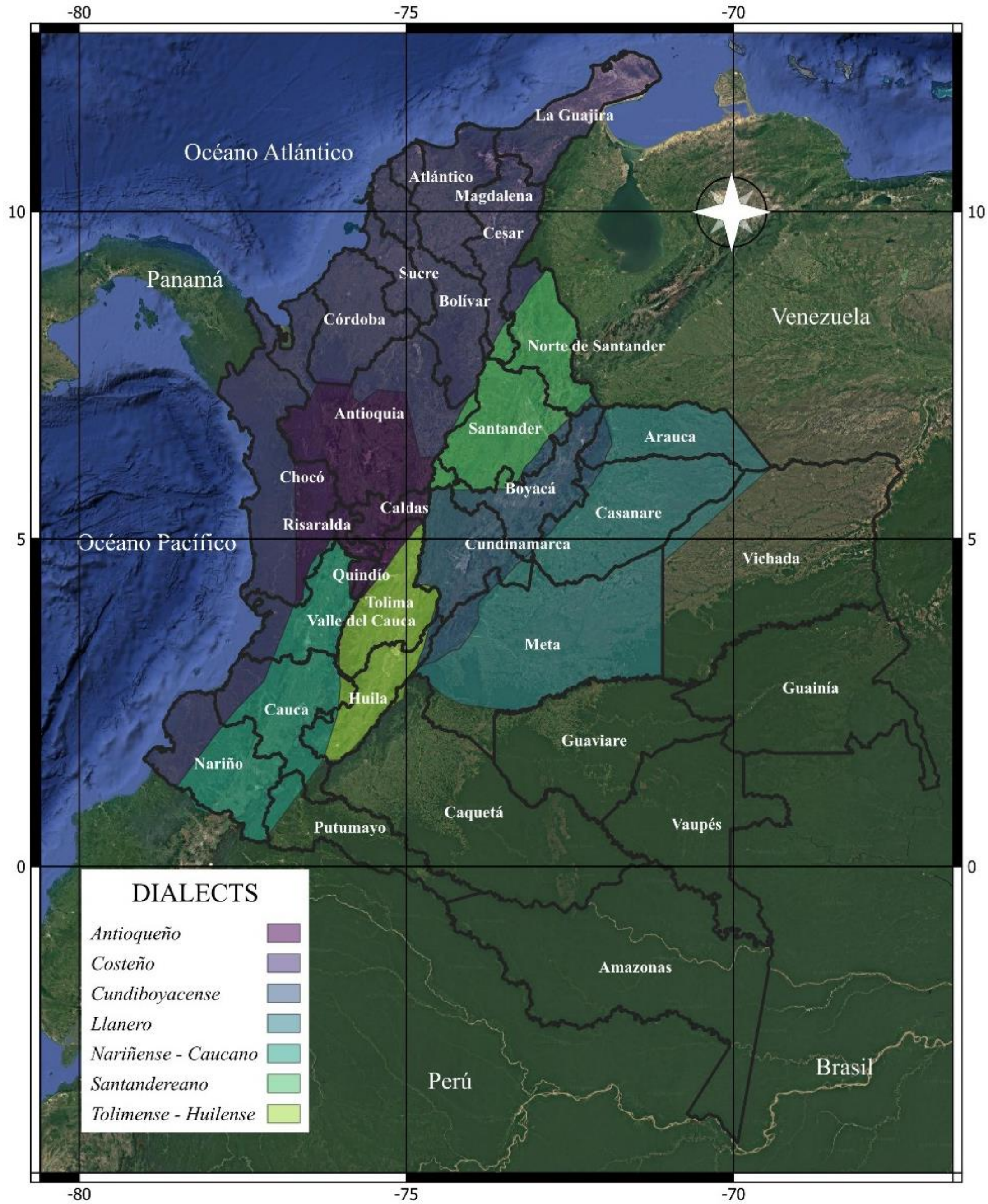
In addition to phonological studies, recent work has addressed lexical variation (Lancheros Redondo, 2018) and emerging grammatical descriptions (Bernal-Chávez & Díaz-Romero, 2017), reflecting the growing diversity of topics in the study of Colombian Spanish. As Markič (2017, p. 186) observes, Colombian Spanish constitutes a national variety composed of a range of regional accents, each with distinct features of pronunciation, intonation, and vocabulary. These internal distinctions contribute to its differentiation from other Spanish varieties spoken across Latin America.

2.1.1 The Dialects of Colombian Spanish after Flórez (1961) and Montes (1982)

One of the country's most significant linguistic contributions is *Atlas Lingüístico-Etnográfico de Colombia* (The Colombian Linguistic-Ethnographic Atlas), widely known by its Spanish acronym, ALEC (Instituto Caro y Cuervo, 1981). This was the first attempt to describe the features that make Colombian Spanish different and was a pioneer project in the study of the Spanish varieties spoken in Latin-America. The research was carried out from 1955 to 1983, time during which investigators collected speech samples from different regions of the country. This collection served as the base for the first dialectal description of the Spanish of Colombia by Flórez (1961). Flórez's account divided Colombia's most densely populated regions into seven dialectal areas: the Costeño, Antioqueño, Nariñense-Caucano, Santandereano, Cundiboyacense, Tolimense-Huilense, and Llanero dialects. Figure 1 presents the geographical distribution of these dialects.

Figure 1

Geographical distribution of Colombian dialects, based on Flórez (1961).



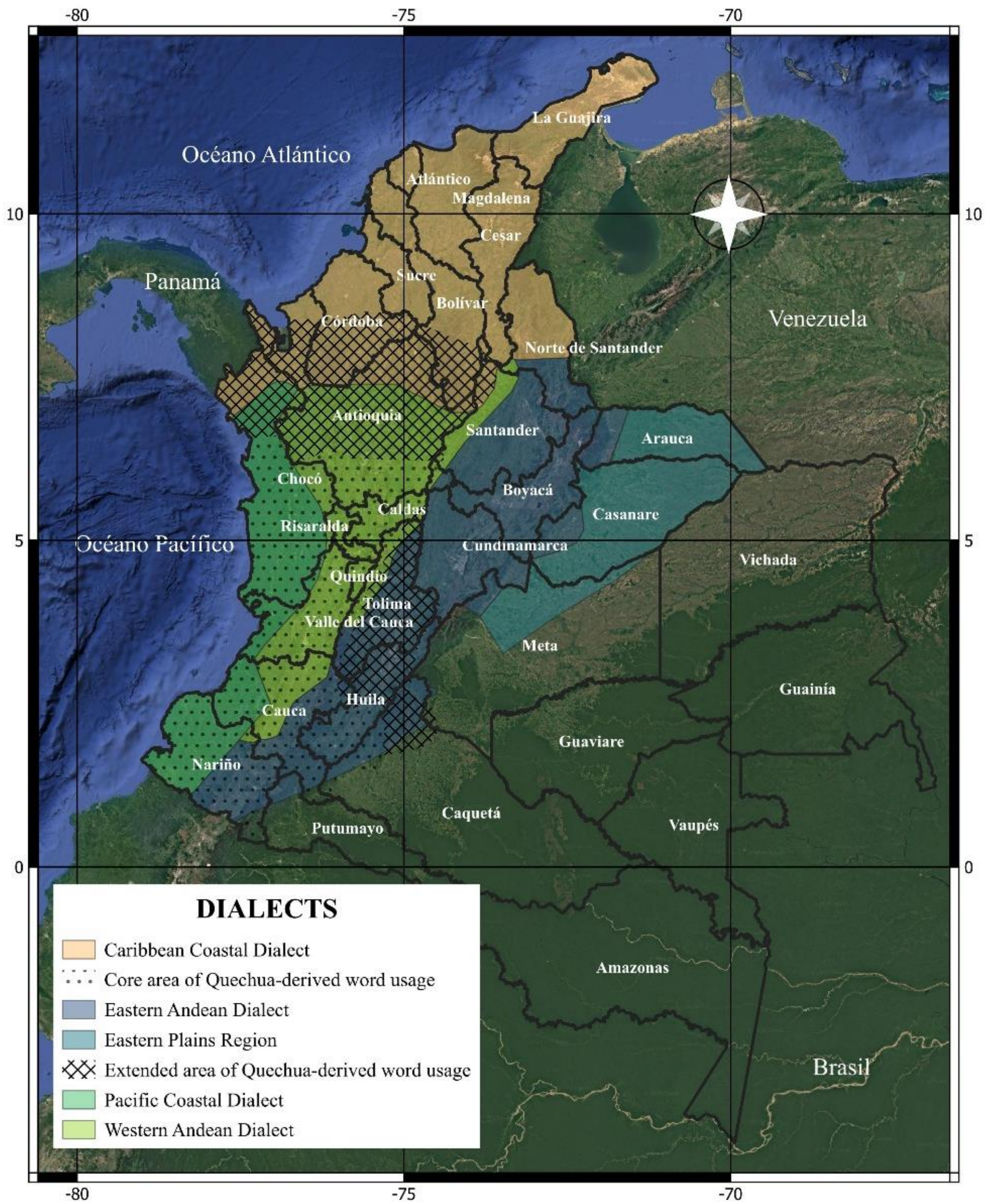
Note. Adapted from 'El español de Colombia. Nueva propuesta de división dialectal' by N. F. Ruiz Vásquez, 2020, *Lenguaje, 48* (2)*,* pp. 160-195.

Following Flórez, Montes (1982) classified Colombian Spanish into macro-dialects, dialects, subdialects, and regional speech varieties. According to Montes (1982), a macro-dialect features several linguistic norms and a dialect presents at least one of these norms. In addition, a subdialect is a further classification of each dialect that can be accounted for based on lexical, phonetic or morphosyntactic features. Finally, subdialects can split into regional speech varieties, which, as explained by Ruiz Vásquez (2020), Montes classified on the basis of lexical differences. According to (Montes, 1982), Colombian Spanish is organized into two macro-dialectal regions: the Coastal Macro-dialect and the Andean[7] (or Central) Macro-dialect. Each of these is further divided into several dialects, which I will outline below, and which are illustrated in Figure 2.

[7] The Andean Macro-Dialect is also known as the Central Macro-Dialect given that the Andean Mountain range falls within the centre of the country. For the purpose of this work, I will refer to it only as Andean Macro-Dialect

Figure 2

Geographical distribution of Colombian dialects, based on Montes (1982).



Note. Adapted from 'El español de Colombia. Nueva propuesta de división dialectal' by N. F. Ruiz Vásquez, 2020, *Lenguaje, 48* (2), pp. 160-195.

According to Montes, the Coastal Macro-dialect is composed of two main dialects: the Pacific Coastal Dialect and the Caribbean Coastal Dialect. These encompass the territories along Colombia's Pacific and Caribbean shores, extending inland until the central Andean region begins, the domain of the Andean Macro-dialect. However, Montes (1982) provides a detailed subdivision only for the Caribbean Coastal Dialect, noting the lack of sufficient data on the dialectal configuration of the Pacific coast. Within Caribbean Spanish, he distinguishes four subdialects: Cartagenero, Samario, Guajiro, and Interior. The Andean Macro-dialect is classified into two primary dialects: the Western Andean Dialect and the Eastern Andean Dialect. These are further divided into subdialectal regions. The Western Andean Dialect includes the speech varieties of the departments[8] of Nariño, Antioquia, and Cauca, while the Eastern Andean Dialect encompasses those of Tolima, Huila, Santander, Cundinamarca, and Boyacá. The latter two departments form what is commonly referred to as the Cundiboyacense Plateau, from which the Cundiboyacense subdialect arises.

2.1.2 Ruiz Vásquez's (2020) Dialectal Scale Model

While the contributions of both Flórez and Montes remain significant, a key limitation of their proposals is the exclusion of the Amazonía and Orinoquía regions. This omission was largely due to factors such as geographic inaccessibility, low population density, and the complex demographic composition of these areas. According to Ruiz Vásquez (2020, p. 178), Montes' definition of macro-dialect would benefit from further refinement, particularly with regard to its

[8] Colombia's political and geographic organization is based on departments. The country comprises 32 departments, which include both continental and insular territories.

geographical scope and the internal consistency of its configuration as a dialectal unit. Consequently, Ruiz Vásquez proposes a new model for classifying dialects more systematically.

His DIALECTAL SCALE MODEL (Ruiz Vásquez, 2020) is based on the percentages of occupied territory by each of the dialectal units, ordered from biggest to smallest; as follows: A MACRO-DIALECT corresponds to a transnational or national unit that covers from 50% to 100% of the space of the studied territory; a DIALECT covers between 12,5% to 50%; a SUBDIALECT covers between 3,12 and 12,5%; and, finally, a LOCAL SPEECH variety corresponds to a scale between 1% to 3,12% of territory. Ruiz Vásquez's model is of interest, not only for the Spanish of Colombia, but for other varieties within the region. It offers a scalable and comparative framework that situates Colombian Spanish more clearly within the broader linguistic landscape of the region. Table 1 presents Ruiz Vásquez's dialectal units and his Dialectal Scale Model definition and classification in more detail.

While Ruiz Vásquez's model maintains crucial elements of that of Monte's, it offers a more inclusive interpretation of Colombian Spanish since it includes new subdialects within his description that had been previously unaccounted for. In a new, post-conflict scenario, and drawing on Ruiz Vásquez's Dialectal Scale Model, I used this framework to define the boundaries of the UCSI Corpus. In doing so, I aim to contribute to the refinement of Ruiz Vásquez's proposal by providing new empirical data that can help describe previously undocumented linguistic phenomena and either reinforce or challenge existing theories about Colombian Spanish and its dialects.

Table 1

Dialectal Scale Model

Level	Normative and spatial structure
Macro-dialect	Diatopic variety characterised by a set of phonetic norms (although not exclusively). Usually transnational or national (Higher than 50 % of the nation's territory)
Dialect	Diatopic variety characterised by at least one phonetic, morphosyntactic or lexical norm. Usually regional (between 12,5 and 50 % of the nation's territory)
Subdialect	Diatopic variety characterised by at least one phonetic, morphosyntactic or lexical norm. Usually local (between 3,12 and 12,5 % of the nation's territory)
Regional Speech	Diatopic variety characterised by at least one phonetic, morphosyntactic or lexical norm. Usually intra-local (between 1 and % of the nation's territory)
Local Speech	Diatopic variety characterised by at least one phonetic, morphosyntactic or lexical norm. It is usually of small range (to the level of a population nucleus)

Note. Retrieved from 'El español de Colombia. Nueva propuesta de división dialectal' by N. F. Ruiz Vásquez, 2020, *Lenguaje, 48* (2), pp. 160-195.

2.1.3 Colombian Spanish Macro-dialects, Dialects and Subdialects after Ruiz Vásquez (2020)

Ruiz Vásquez (2020)'s proposal consists of three Macro-dialects, from which five dialects of Colombian Spanish can be derived (Table 2 and Figure 3, below illustrate these). Firstly, two of the Spanish macro-dialects included are transnational in scope, whereas the third, newly proposed macro-dialect pertains exclusively to the Colombian national context. The transnational macro-dialects are the Antillean Macro-Dialect and the Andean Macro-Dialect. On the one hand, the Antillean Spanish covers the area corresponding to the Colombian Caribbean Dialect and connects it with Panamá, Venezuela and the Antilles, mainly. Ruiz Vásquez (2020, p. 182) explains that it is defined by pronunciation features as well as morphosyntactic and lexical phenomena shared with the Antilles. On the other hand, the Andean Macro-dialect in Colombia is influenced by the linguistic contact with the Quechua at the Andes. This macro-dialect is shared with the Argentinian Northeast, the Andes of Bolivia, Ecuador and Peru. In Colombia, it is present at the South in the Putumayo Department, up to the North of the country and through the regions over which the Andes Mountain range extends and splits in three, originating valleys between the three mountain ranges (Ruiz Vásquez, 2020, p. 182)

The third macro-dialect is exclusive to Colombia, but it can still be considered *macro* since it covers approximately 50% of the national territory. Ruiz Vásquez (2020, p. 183) names it Neogranadino[9] Macro-Dialect based on historical grounds. It lies between the previously described Antillean and Andean Macro-dialects. Geographically, it can be located within the interior area of the country, which includes the departments that border the Pacific Ocean, Cauca, Chocó, Nariño, and Valle del Cauca. It also extends across the inter-Andean valleys of the Cauca and Magdalena rivers, and, finally, into the Colombian Amazonía and Orinoquía regions, which had not been previously described in earlier classifications. This final macro-

---

[9] Neogranadino comes from the Spanish name *República de la Nueva Granada*, which was the name given to the first republic that constituted most of the Colombian present territories.

dialect was described based on phonetic information related to the realizations of the fricative /s/, information that was extracted from the work of Brown and Brown (2012) and Ramírez and Almira Vázquez (2016)

Now that the three Macro-Dialects, Antiallean, Andean and Neogranadino have been introduced, Table 2 and Figure 3 present how they further split into dialects, subdialects and regional speech varieties, as proposed by Ruiz Vásquez (2020).

Table 2

New dialectal division for the Spanish of Colombia (Ruiz Vásquez, 2020)

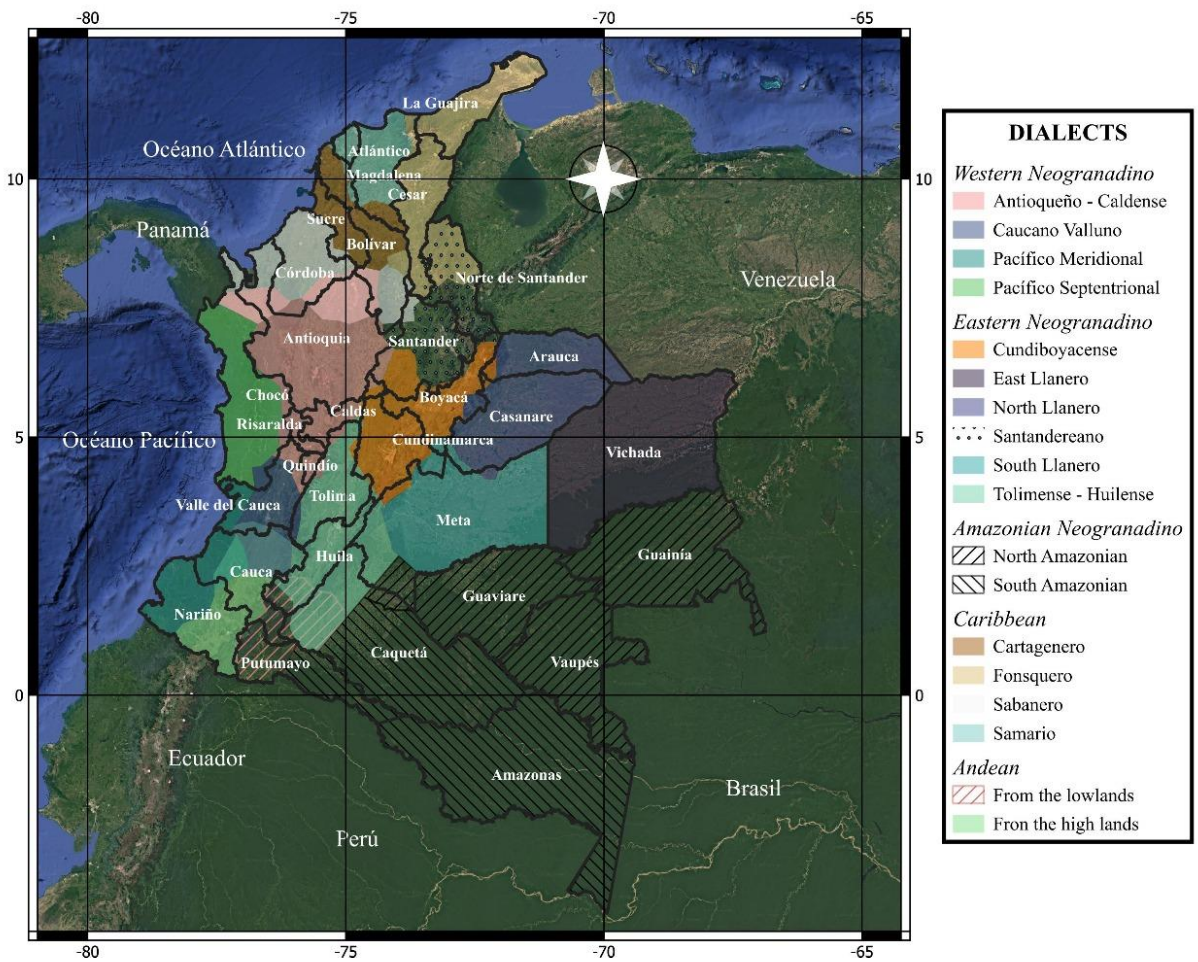
Macro-Dialect	Dialect	Subdialect	Regional Speeches
Antillean Spanish	Colombian Caribbean	Cartagenero	North Bolivar / Sucre
		Samario	Atlántico / Magdalena
		Fonsequero	Guajira / Cesar / North Norte de Santander
		Sabanero	Córdoba / Urabá Antioqueño / South Bolívar
Neogranadino Spanish	Western	Antioqueño-Caldense	Antioquia / Caldas / Quindío Risaralda / North Valle del Cauca
		Caucano-Valluno	Cauca Andes / The valley of Cauca
		Pacífico Septentrional	Chocó
		Pacífico Meridional	The coast of Valle del Cauca / Cauca / Nariño
	Eastern	Santandereano	Santander / South Norte de Santander

		Cundiboyacense	Boyacá / Cundinamarca
		Tolimense-Huilense	Tolima / Huila / Western Caquetá
		North Llanero	Arauca / Casanare
		South Llanero	Meta
		East Llanero	Vichada
	Amazonian	North Amazonian	Guaviare / Guianía / Vaupés
		South Amazonian	East Caquetá / Amazonas
Andean Spanish	Colombian Andean	From the highlands	The Andes of Nariño
		From the lowlands	Putumayo

Note. Adapted from ‘El español de Colombia. Nueva propuesta de división dialectal’ by N. F. Ruiz Vásquez, 2020, *Lenguaje, 48* (2), pp. 160-195.

Figure 3

Geographical illustration of the New dialectal division for the Spanish of Colombia based on Ruiz Vásquez (2020).



Note. Adapted from 'El español de Colombia. Nueva propuesta de división dialectal' by N.F. Ruiz Vásquez, 2020, *Lenguaje*, *48* (2), pp. 160-195.

Ruiz Vásquez explains that for East Llanero (Eastern Neogranadino) and North and South Amazonian (Amazonian Neogranadino) there are no sufficient studies that can corroborate this dialectal subdivision (Ruiz Vásquez, 2020, p. 188). In consequence, these three subdivisions remain hypothetical. As previously noted, the UCSI Corpus was designed following the dialectal classification proposed by Ruiz Vásquez. This framework not only aligns with a contemporary view of the dialectal structure in Colombia but also creates opportunities for future research to evaluate, refine, or expand upon current classifications. The data it provides may serve as a foundation for exploring linguistic variation across Colombian Spanish dialects and for empirically testing subdivisions that remain under discussion.

2.2 Sample Design and Size

For the UCSI Corpus, I aimed to collect unscripted speech from speakers of the five major dialects of Colombian Spanish: the Colombian Caribbean, Colombian Andean, Colombian Amazonian, and the Western and Eastern Neogranadino dialects. The goal for the corpus is to record a total of 54 speakers, interacting in pairs, with sampling guided by the aim of covering both macro-dialectal and subdialectal variation. In this thesis, I report on the first phase of data collection in Colombia, which includes recordings from 38 participants, 2 of whom were used for piloting, resulting in a core dataset of 36 speakers. This number was defined in line with the timeline and scope of the doctoral research.

During sampling, I prioritized comprehensive representation of the five major dialects over full balance across subdialects and regional speech varieties. Nonetheless, I aimed to include at least one pair of speakers from each subdialect within the broader dialect zones. As a result, the number of speakers per dialect varies, reflecting internal dialectal complexity. For instance, the Eastern Neogranadino dialect comprises six subdialects, and therefore was allocated a higher target of 18 speakers. In contrast, the Neogranadino Amazonian dialect is divided into only two subdialects, and correspondingly, six speakers were recruited from that region.

Although this thesis does not focus on the role of social variables in the functioning of repair strategies, a set of sociolinguistic variables was included during corpus design. These were kept as balanced as possible across speakers, in order to support future research using the UCSI Corpus that may wish to explore the influence of social factors, whether in repair or in other linguistic phenomena.

To that end, I sought to pair participants by age, social class, and self-identified gender, as far as was feasible. For the variable AGE, three groups were defined covering the following ages: 18 to 35 years old, 36 to 55 years old; and 56 to 70 years old (See Table 3). Regarding social class, legislation in Colombia defines six socio-economic levels to stratify the urban areas of the country (DANE, 2015; Senado de la República, 1994). Under these regulations, both the national and the local governments establish the rate of public services and the granting of subsidies for the general population. Following this system, I grouped participants into three broader categories: levels 1, 2, and 3 as low, level 4 as middle, and levels 5 and 6 as high. Table 4 illustrates this classification.

Table 3

Age groups

Group	**Age range**
1	18 to 35 years old
2	36 to 55 years old
3	56 to 70 years old

Table 4

Social and economic stratification in Colombia

Level	Stratification
1	Low-low
2	Low
3	Mid-low
4	Mid
5	Mid-high
6	High

Note. Adapted from *Metodología de estratificación socioeconómica urbana para servicios públicos domiciliarios* by Departamento Administrativo Nacional de Estadística (DANE) (2015), and *Ley 142 del 11 de julio de 1994* by Senado de la República (1994).

Lastly, the following aspects were also considered. Participants should not have lived for more than six months in a different location from the ones corresponding to each dialect and subdialect in the last two years. In addition, they should report not having or having had any speech or hearing disorders. See Table 5 for details on the final dialectal classification and social stratification of the UCSI corpus.

Once the sample design was established, different people from all regions of the country were recruited by the researcher, either by email or text. Each participant was asked to invite a friend with a similar sociolinguistic background who would be willing to participate and interact with them during the recording. After being informed about the aims of the project and providing their consent, participants were invited to attend the premises of the Universidad Nacional de Colombia in Bogotá for the recordings.

Table 5

UCSI Corpus: Dialectal and social stratification

Dialect	Pair	Subdialect	Gender	Age group	Social stratification
Colombian Caribbean	1	Cartagenero	F	1	L
	2	Samario	M	2	M
	3	Fonsequero	F	3	H
	4	Sabanero	M	1	L
	5	To be determined	F	2	M
	6	To be determined	M	3	H
Western Neogranadino	7	Antioqueño-caldense	F	1	L
	9	Caucano-Valluno	M	2	M
	10	Pacífico Septentrional	F	3	H
	11	Pacífico Meridional	M	1	L
	12	To be determined	F	2	M
	13	To be determined	M	3	H
Eastern Neogranadino	14	Santandereano	F	1	L
	15	Cundiboyacense	M	2	M
	16	Tolimense-huilense	F	3	H
	17	North Llanero	M	1	L
	18	South Llanero	F	2	M
	19	East Llanero	M	3	H
	20	To be determined	F	1	L
	21	To be determined	M	2	M
	22	To be determined	F	3	H
Colombian Amazonian	23	North Amazonian	M	1	L
	24	South Amazonian	F	2	M
	25	To be determined	M	3	H
Colombian Andean	26	From the highlands	F	1	L
	27	From the lowlands	M	2	M
	28	To be determined	F	3	H

2.3 Materials

This section describes the materials used to elicit the unscripted speech data that constitute the core of this project and the UCSI Corpus. First, the Diapix and Consensual Response tasks will be presented. Then, I will talk about the language background questionnaire that was implemented to gather other relevant sociolinguistic information for profiling participants.

2.3.1 The Diapix Task

The Diapix task was designed for eliciting unscripted interactive linguistic data (Baker & Hazan, 2011; Tuomainen et al., 2022; Van Engen et al., 2010). The tasks consist of a 'spot the difference' game in which participants need to find the differences between two images. The task was performed in pairs, with no intervention from the researcher during the conversation. In each set of images there are at least 12 differences. The objective for both speakers was to identify those differences, as fast as possible, by describing the pictures verbally and without looking at each other's image. The sets of pictures were distributed in three semantic domains: *Beach*, *Farm* and *Street*. Since the introduction of the first Diapix, which was designed for English (Van Engen et al., 2010), there have been different adaptations of it, including one for British English (DiapixUK) (Baker & Hazan, 2011).

One of the most recently developed ones was the one proposed by Figueroa et al. (2019) for Spanish (DiapixSP). This task is a full adaptation of the DiapixUK protocol, developed with the aim of eliciting spontaneous speech from speakers of South American Spanish varieties. The adaptation involved a comprehensive cultural and linguistic revision of 12 pairs of images (plus one practice pair) from the original DiapixUK materials, resulting in a total of 197 modifications. Most changes consisted of non-literal translations and the avoidance of region-specific colloquialisms, in order to ensure that the images could be used across a range of Hispanic contexts. DiapixSP presents a total of 12 pairs of images, all in Spanish, with four pairs for each semantic domain. It also includes one pair from the *Farm* domain that is used for practice, for a total of 13 pairs. Similarly to what happens with the DiapixUK, each pair of images includes 12 differences distributed homogeneously. It is assumed that each pair has a similar level of difficulty. Finally, while the original DiapixUK materials included certain phoneme contrasts embedded within the images, this feature was not retained in the DiapixSP version. As reported by Baker and Hazan (2011) and discussed by (Figueroa et al., 2019, p. 270), including targeted phonemic contrasts does not appear to significantly increase the likelihood that speakers will produce the expected phonemes during spontaneous speech. Figure 4 presents an example of the adaptation from DiapixUK to DiapixSP.

To this project, the whole adaptation of the DiapixSP was adopted and no major changes were made to the images. The one exception was the set of pictures belonging to the *Beach* semantic domain in which a taxi for the DiapixSP was coloured in black, as in Argentina and Chile; taxis in Colombia are mainly yellow so that change was incorporated.

Figure 4

Comparison between the DiapixSP (Figueroa et al., 2019)(top) and the present project (bottom).

Note. Adapted from 'DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs' by R. Baker and V. Hazan, 2011, *Behavior research methods* 43(3), pp. 761-770 and 'DiapixSp: adaptación al español y aplicación piloto de una herramienta de elicitación de habla espontánea y colaborativa' by M. A. Figueroa, D. A. García, and G. F. Salamanca, 2019, *Estudios de fonética experimental* (28), pp. 257-288.

2.3.2 Consensual Response Task

The second task for eliciting unscripted speech involved a dialogue in which participants were required to agree on responses to two questions and report them in writing. To that end, they were presented with five different questions on varied topics, each accompanied by a brief written contextualization. The topics were carefully selected by the researcher with the goal of prompting emotionally engaged and spontaneous language use. These included controversial or personally sensitive issues designed to elicit responses in which speakers were less likely to monitor or control their speech closely. This approach was inspired by Labov's (1972) 'danger of death' question, which highlights the value of emotionally charged topics in reducing speakers' self-consciousness and eliciting more naturalistic discourse. In addition, the questions were phrased to avoid simple *yes-or-no* answers, so that participants were invited to take a position, discuss and explain their choices. Table 6 illustrates one of the topics that were used to elicit speech in this task. Appendix C presents the all the options that were offered to participants, both in Spanish and English.

To complete the task, participants were invited to take some time to read through the five topics and then select at least two questions based on their preferences. They were also provided with a blank sheet to write down their agreed-upon responses. Only one participant assumed the role of writer. While working on the written portion, participants were expected to interact and engage in discussion. No specific instructions were given regarding length or writing style, as the primary aim was to prompt spontaneous dialogue as part of the interaction. Participants were given a total of 18 minutes to complete the task. Once the instructions were delivered and the questions selected, I left the room, allowing participants to discuss, reach a consensus, and compose their written response.

Table 6

English translation illustrating one topic and its question

Topic	Contextualization	Question
Life sentence in Colombia	Recently, the Constitutional Court[10] stated that the political reform punishing child rapists with a life sentence was unconstitutional. The latter happened even though the proposal had been well-received by some sectors, given the constant number of rape cases that are being reported recently.	Do you consider that the Constitutional Court made the right decision? Or, quite the opposite, do you believe that the above-mentioned reform should be reconsidered and implemented, for this and even other crimes?

2.3.3 Language Background Questionnaire

The main objective of the language background questionnaire was to collect detailed information to confirm participants' dialects and subdialects, as well as their age, socio-economic background, and knowledge of other languages for future projects. The questionnaire gathered information not only about the participants themselves but also about their parents or primary caregivers, given the potential influence of family background on speech patterns. This questionnaire also served to confirm whether participants had speech, hearing, or other health conditions that could affect their participation. For further details, both the English and Spanish versions of the Language Background Questionnaire are included in Appendix A.

[10] The Constitutional Court is the national entity in charge of protecting the integrity of the constitution and the democracy in Colombia.

2.4 Ethical Considerations

To gain ethical clearance for this PhD project, I needed to comply with the requirements for research involving human participants in Colombia, where the fieldwork portion of this project took place as well as with rules and regulations from the United Kingdom and the University of Leeds. This section briefly outlines the approval process. For further details, Appendix B shows all ethical aspects that were considered for developing this PhD thesis, including information on informants' anonymity, data collection and safeguarding, and the specifics of research approval from both institutions involved.

This research was submitted as an independent project at the Universidad Nacional de Colombia (the principal investigator's home institution). The project was approved by the University and granted ethical approval under Colombian law (Resolución 008430 de 1993 [Resolution 008430 from 1993]) governing research with human participants. According to Article 11 of the above-mentioned Resolution, this research was considered as 'low risk' given that the data collected consisted of documents and no underlying physical, biological or psychological conditions of participants were modified. In addition, biosecurity and health protocols related to the COVID-19 pandemic were also strictly observed. Ethical approval was granted by the Faculty of Human Sciences following review by the Head of the Linguistics Department.

Once ethical clearance was granted in Colombia, I applied to the University of Leeds Research Ethics Committee. Apart from the information presented earlier, this application also reported on details regarding data management and included an English translation of the Consent Form and Information Sheet given to participants (Appendix B). The Arts, Humanities and Cultures (AHC) Research Ethics Committee approved the project in September 2021 (reference number, FAHC 20-102).

2.5 Procedure

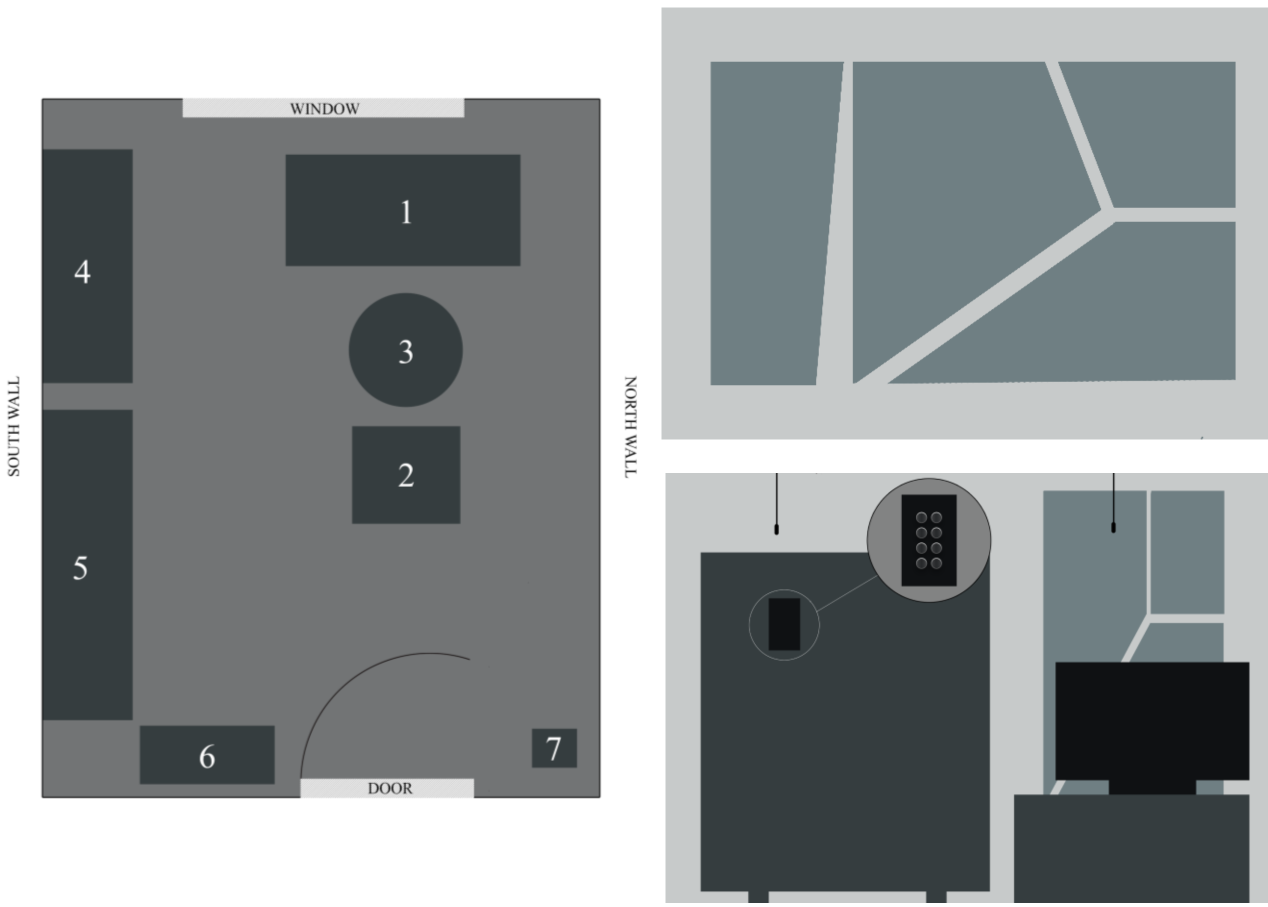
In this section I describe the procedure followed for data collection. First, I will outline the setting and equipment used, followed by a step-by-step description of each data collection session. Recordings for this project took place at the Universidad Nacional de Colombia (Bogotá Campus).

2.5.1 The Interactional Linguistics Unit at Universidad Nacional de Colombia

At the Department of Linguistics, and within the Laboratory of Linguistics, there is an INTERACTIONAL LINGUISTICS UNIT, which is equipped for high-quality video and audio recording. The Unit is composed of a sound-treated room, THE SUITE, and a CONTROL ROOM. This space was designed so that speakers can freely interact, mitigating the feeling of being recorded and known within the field of sociolinguistics as the Observer's Paradox (Labov, 1972). The Suite within the Unit does not resemble a usual laboratory setting as the space has been arranged to look like a living room. The arrangements aimed to foster a comfortable environment, helping speakers feel less conscious of being observed or recorded. Within the Unit, efforts have been made to keep audio and video instruments out of sight of participants. Figure 5 illustrates the staging of the Suite at the Interactional Linguistic Unit.

Figure 5

Layout and interior view of the recording suite at the Interactional Linguistics Unit, Department of Linguistics, Universidad Nacional de Colombia.



Note. The *Left image* shows a top view of The Suite; which includes: 1. A sofa; 2. An armchair; 3. A coffee table; 4. A TV table and TV; 5. A library; 6. An auxiliary table; and 7. A coat rack. The *top right image* shows the North wall of the room, which is fitted with sound absorption panels; these acoustic panels were specially designed for the space and are in all walls and ceiling. The *bottom right image* displays the South wall in which the audio patch panel is located, hidden by the library; next to it there is the TV and TV table; also in this image, two air microphones are visible.

As mentioned earlier, the room itself has been sound-attenuated. All walls and ceiling were treated with Dense Fibre Matting slabs. Both the door and window are soundproof. In addition, walls and ceiling were fitted with sound absorption panels. Altogether, these actions prevent the transmission of sound into the room, creating an ideal space for the recording of high-quality speech materials. The suite is connected to the adjacent control room via an audio patch panel. Recordings were managed from the control room using a DELL Precision M2800 desktop, which was connected to a PreSonus Studio 1824C audio interface. This interface supports up to eight audio inputs and two outputs.

2.5.2 Recordings

For the recordings conducted for this project, all data collection sessions were led by me, the author, a native speaker of Colombian Spanish. Each pair of participants completed all tasks in one unique session. The following section outlines the organization of these sessions and describes the detailed procedure that was implemented.

As soon as participants arrived in the Laboratory, they were asked to read the information sheet, and after agreeing to take part in the study, sign the consent form. Secondly, informants were asked to remove their facemasks and were provided with new ones. The latter was decided given that to the COVID-19 pandemic at the time, face covering was required in public spaces in Colombia[11]. Following this biosecurity protocol, the researcher again explained the step-by-step

[11] An impressionistic comparison of waveforms and spectrograms from the pilot recordings indicated that the use of face coverings did not significantly affect the sound quality. However, to control for any potential effects and to ensure consistency across recordings, all participants were asked to wear the same type of face mask—a NexCare 3M Ear Loop Face Mask, known for being lightweight and breathable, whenever face coverings were required.

procedure and informed participants that the whole session was planned to last approximately one hour. Subsequently, participants were equipped with wireless head-mounted Shure BLX14-P31-H9 microphones, and the different tasks and recordings began. Each participant was recorded on a separate channel using the PreSonus audio interface and the software Audition (Adobe Inc., 2019).

As a first task, participants were asked to read the passage *The North wind and the Sun* [El viento norte y el sol][12] (Coloma, 2016, 2018). This part of the session helped the me adjust the gain to level microphone signals and confirm that facemasks were not in contact with the microphone, which would cause noise to the signal.

Next, the Diapix task was introduced. A practice round, lasting approximately four to five minutes, was conducted first. After this period, I returned to the suite, paused the task, checked in with participants about their experience, and addressed any questions or uncertainties regarding the procedure. Once the instructions had been clarified and the participants felt confident, the experimental run of the Diapix task began.

Each pair of participants completed two sets of images. On average, it took approximately 12 minutes per set for participants to identify at least 10 differences. In total, the Diapix task recordings lasted about 30 minutes. Most pairs completed each round in 10 to 15 minutes; however, a few took longer, and in those cases, the recordings were allowed to continue without interruption. Two pairs completed the task in less than eight minutes per round. To ensure a

[12] Coloma (2016) proposes a new version of this phonetically balanced text for Spanish. The proposal includes two extra phonemes, less repetition of words and a more balanced distribution in comparison to previous versions by Martínez-Celdrán et al. (2003) and Monroy & Hernández-Campoy (2015).

minimum of 20 minutes of speech was recorded from each pair, these participants were given a third set of images and completed an additional round of the task.

After the Diapix conversations were finished, participants were given a short break of approximately 10 minutes. Subsequently, the consensual response task started. I explained the tasks and awaited while the participants selected the topic of their choice. Once instructions were confirmed, the researcher again exited the room and left participants to carry out the conversation independently. During the discussion, I monitored the session from the control room to check whether the participants had reached an agreement on the selected questions. If required, I returned briefly to assist the participants in selecting a new topic in cases where no consensus had been reached within the first 12 minutes. This task took 20 minutes, on average.

Once the consensual response task was recorded, participants were informed that the sessions was finished, and microphones were removed. To conclude the session, participants were asked to complete the Language Background Questionnaire. This form was administered only after the speech tasks to avoid influencing participants' linguistic performance. Because the questionnaire includes detailed questions about language background and social factors, there was a risk that completing it beforehand could have led participants to adjust their speech, either consciously or unconsciously, to align with perceived expectations (e.g., by adopting a more formal or 'prestigious' variety). This concern reflects Labov's (1972) observer's paradox, which highlights the difficulty of eliciting natural speech when participants are aware they are being studied.

2.6 Data Processing

This section outlines the procedures for data management and processing. I begin by describing how the data were organized, stored, and safeguarded throughout the project in accordance with ethical and institutional guidelines. I then detail the transcription protocols applied to the corpus, including the different layers and levels of annotation used to prepare the data for analysis. Where relevant, examples are provided to illustrate the structure and scope of the transcriptions applied to the UCSI corpus.

2.6.1 Data Management

In accordance with the designed data management plan, all recordings were pseudo-anonymized, and each participant was assigned a unique code (e.g., 001, 002, etc.). Recordings were later organized and labelled following a standardized naming convention that includes, project acronym, participants' code, task type and run and researcher's initials. Given that each recording lasts from 15 to 20 minutes (approximately), all files were segmented into one-minute fragments to facilitate handling and processing. In consequence, each file name follows a consistent structure, as shown below:

sp	001 002	dpix	1	kv	part5
project name	*participants*	*task*	*run*	*researcher initials*	*audio file number*

This naming convention should be interpreted as follows: *project name* refers to Spanish Repair (sp). *Participants* were given ordinal codes. The *task* refers to either the Diapix (dpix) task, the Consensual Response (CR) or the reading exercise (read). Only the Diapix task will also include a column for *run;* 0, for the practice run; 1, for the first sets of pictures, and so on. This systematic naming structure ensures that each audio fragment can be easily identified, retrieved, and cross-referenced, thereby enhancing the reproducibility of the study and supporting future analyses using the UCSI Corpus.

Finally, regarding data management, data was collected using a password-protected laptop to then be transferred to the University of Leeds *OneDrive* as soon as was practical. Data in the form of recordings, orthographic transcriptions and annotations were deposited at the *Research Data Leeds Repository*, in accordance with its data deposit guidelines. Lastly, all materials derived from the UCSU Corpus used in this project, including annotated repair sequences and timing analyses, have also been made publicly available through the Open Science Framework (OSF) Repository at https://osf.io/zrqxd/, supporting open science practices and ensuring the transparency and reproducibility of this research.

2.6.2 Transcriptions, Annotations, and Repair Collections

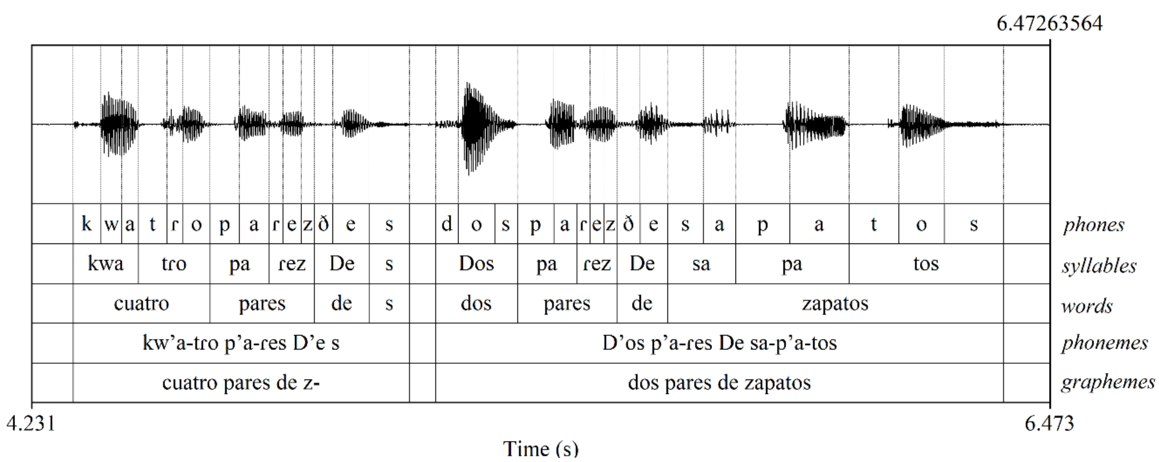
The UCSI Corpus was transcribed at two levels, depending on its intended use. For the full corpus, orthographic transcriptions of the one-minute audio fragments were generated using Microsoft's Azure Speech Service (Batch Transcription), with Spanish as the target language. These transcriptions were reviewed and lightly corrected by the researcher to ensure consistency and alignment with the project's transcription conventions.

A subset of the corpus containing lexical self-repairs was extracted for more detailed phonetic and phonological analysis. For these fragments, the transcription process was carried out entirely within Praat[13] (Boersma & Weenink, 2021). Orthographic transcriptions were manually created by the researcher using the TextGrid function, and were subsequently used as input for a semi-automatic alignment process in EasyAlign (Goldman, 2011) with the Spanish extension developed by (Goldman & Schwab, 2014).

Each annotated file includes multiple tiers, comprising orthographic (graphemes), phonological (phonemes), and phonetic (phones) transcriptions, as well as segmentation into words and syllables. All tiers were manually reviewed and corrected by the researcher to ensure accuracy. Figure 6 illustrates this transcription procedure and its output.

Figure 6

Resulting five-tiered Praat TextGrid.



Note. Segmented waveform and TextGrid of the repair ‘*Dos pares de zapatos*’ (Two pairs of shoes).

---

[13] *Praat* is a free software tool for the analysis of speech in phonetics, available at www.praat.org.

Lastly, for the prosodic and temporal analyses, additional processing steps were carried out for each repair sequence; these are described in detail in the respective methodological sections of each chapter.

2.7 Corpus Evaluation

The UCSI Corpus was designed to meet four core criteria: to capture spontaneous speech, to provide dialectal representativeness, to include a sufficient number of self-initiated repair sequences, and to offer enough speaker-specific data for studying variation. The following evaluation demonstrates how each of these goals was achieved.

2.7.1 Spontaneous, Unscripted Speech

All interactions in the UCSI Corpus were elicited using tasks designed to prompt natural dialogue between familiar speakers, without researcher intervention during the recording sessions. The conversational nature of the Diapix and Consensual Response tasks, combined with a relaxed recording environment, helped encourage spontaneous, unscripted language use. This condition was upheld across all sessions.

2.7.2 Regional representativeness

The corpus includes data from speakers representing the five major dialectal regions of Colombian Spanish: the Caribbean, Andean, Amazonian, Western Neogranadino, and Eastern Neogranadino dialect zones. Within these, sampling aimed to capture at least one pair of speakers from each major subdialect, ensuring both geographic breadth and internal dialectal diversity. The resulting dataset reflects Colombia's internal linguistic variation and offers opportunities for both regional and cross-dialectal comparison.

2.7.3 Richness of Repair Data

A central focus of the corpus design was to include a high number of self-initiated repair sequences. In total, the corpus contains 1,057 repair sequences, distributed across all speaker pairs and spanning a range of interactional contexts. The majority of these are *self-initiated, self-repair* sequences (n = 923), which are the primary focus of this thesis. The data set also includes self-initiated, other-repair sequences (n = 67), other-initiated, self-repair (n = 44), and other-initiated, other-repair (n = 23). When combined, self-initiated repairs account for 990 sequences, making self-initiation by far the most frequent repair initiation type in the corpus. This distribution supports both a robust analysis of the dominant repair configuration and the potential for exploratory work on less frequent types. Additional annotations applied to the self-initiated, self-repair sequences are described in detail in the corresponding analytical papers (Chapter 3 on temporal organization and Chapter 4 on prosodic marking).

2.7.4 Speaker Coverage and Future Usability

Each speaker in the corpus contributes a substantial amount of speech, with recording sessions averaging approximately 45 minutes of spoken interaction per pair of speakers. This allows for analyses of inter-speaker variation, as well as within-speaker consistency across different tasks. Beyond the scope of this thesis, the corpus represents a valuable resource for future studies on dialectal variation, prosodic features, temporal organization, and sociolinguistic factors in Colombian Spanish.

Together, these results confirm that the UCSI Corpus fulfils its design objectives and offers a well-documented, dialectally diverse dataset for the study of spontaneous speech and repair in Colombian Spanish.

Chapter 3

Temporal Organization in Lexical Self-repair

3.1 Introduction

This chapter reports on an investigation on the temporal organization of lexical self-repair in unscripted Colombian Spanish extracted from the UCSI corpus, which construction was described in Chapter 2. Analyses were carried out by implementing a Bayesian modelling framework and unsupervised machine learning by clustering.

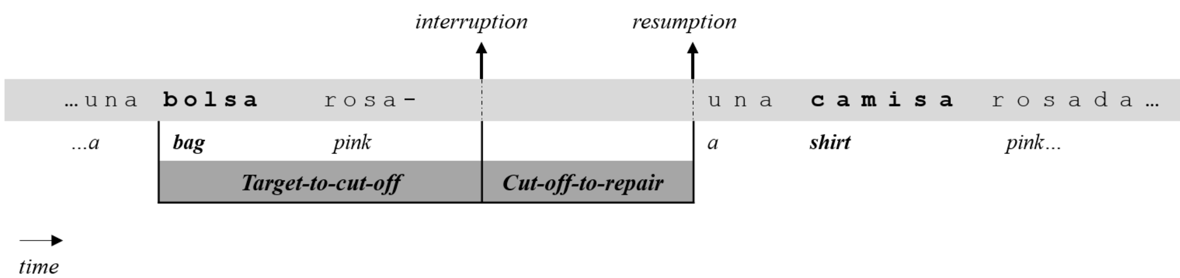
In lexical self-repair, a speaker rejects one lexical choice for another, as illustrated earlier in example (9), which reads *una* ***bolsa*** *rosa-* (...) *una* ***camisa*** *rosada*, in English, 'a pink **bag** (...) a pink **shirt**'. In the example, the noun *bolsa* 'bag' is the reparandum and *camisa* 'shirt' is the repair; both are produced and corrected by the same speaker in the same turn. In the introduction of the thesis (Section 1.3.1) I discussed how research on temporal organization of repair has used the relevant timing intervals around the reparandum and the repair to analyse and describe the strategies used by speakers in different repair sequences. The analyses presented in this chapter build on this work and examine the TARGET-TO-CUT-OFF[14] and the CUT-OFF-TO-REPAIR timing intervals, as illustrated schematically in Figure 7 below, with respect to time. The first interval is the time between the start of the reparandum, or problematic 'target word' attempt, *bolsa* 'bag'

[14] In many of the papers I review in this chapter, authors have named this interval 'Error-to-cutoff'. I have opted for the term 'Target', as in Target-to-cut-off, because this thesis included not only errors but also appropriateness repairs. The term 'Target' to refer to the troublesome item or reparandum, which serves as reference for my measurements of timing, is broader and includes the different subtypes of repair that are relevant to my reports.

in the example, and the cut-off, or interruption; in this case, the moment at which the speaker abandons the articulation of the adjective *rosada* [roˈsa-] ‘pink’. Secondly, the Cut-off-to-repair is the timing interval between the moment of interruption, again at *rosada* [roˈsa-] ‘pink’, and the resumption of talk, which falls at the start of the first word of the repair sentence, *una* ‘a’.

Figure 7

Schematic representation of the Target-to-cut-off and the Cut-off-to-repair intervals.



In Section 1.2.2, I presented the semantic classification between appropriateness repairs and linguistic and factual error repairs; the former, recall, refers to issues of pragmatic felicity, as seen in (14), whereas the latter refer to inaccuracies that are either linguistically incorrect or factually wrong, as shown in (15) and (16), respectively. This chapter examines the characteristics of both the Target-to-cut-off and Cut-off-to-repair intervals, their temporal organization, and how the semantic division between APPROPRIATENESS, FACTUAL, and LINGUISTIC repairs shapes their timing individually and in interaction.

(14) APPROPRIATENESS, lexical, self-initiated, self-repair

Spanish (Vera Diettes, UCSI Corpus, 149, sp_021_022_ conv_kv_part4)

01	A:	Ese **performance** está, como de alguna forma, tan bien montado-; esa [ˈe.sa:] **obra de teatro** está tan bien montada. *That **performance** is, in a way, so well put up-; that **performance** is so well put up.*

(15) FACTUAL ERROR, lexical, self-initiated, self-repair

Spanish (Vera Diettes, UCSI Corpus, 54, sp_009_010_dpix_1_kvpart1)

01	B:	¿Cuántos niños hay cerca a la canasta? *How many children are there near the basket?*
02	A: →	**cuatro**, bueno, **tres**; hay un niño que está cerca pero está mirando hacia otro lado ***four**, well, **three**; there is one kid that is near but is looking to the other side*

(16) LINGUISTIC ERROR, lexical, self-initiated, self-repair

Spanish (Vera Diettes, UCSI Corpus, 146, sp_023_024_conv_kv_part26)

01	A:	En qué **agravante**-; ¿en qué le **empeora** eso la vida a la persona? *How does this **aggravation**-; how does this **worsen** someone's life?*

I also evaluate the effects of the reparandum COMPLETENESS on both intervals to determine whether the distinction between complete and incomplete (i.e., interrupted) target words, as seen in (17), has an impact on the timing and coordination of both Target-to-cut-off and Cut-off-to-repair. Lastly, I examine the presence and absence of EDITING TERMS, interjections such as 'uh' or 'I mean', which are also exemplified in (17).

All things considered, this chapter investigates and reports on the temporal make-up of repair in relation to repair semantics, reparandum completeness, and the presence of editing terms.

(17) Self-initiated repair with an INCOMPLETE reparandum and EDITING TERMS

Spanish (Vera Diettes, UCSI Corpus 382, sp_021_022_dpix1_kv_part1)

01	A:	y un **vestid- pues sí,** como un **traje** naranja con falda *and a **dress**, yes, like a **gown**, orange with a skirt*

By addressing the issues presented above, this chapter contributes to understanding whether previous findings on the phonetic temporal aspects of repair in other languages occur similarly in non-Germanic languages, such as Spanish, and, more specifically, in the varieties spoken in Colombia. This contribution ties directly to the overarching aim of the dissertation, which is to determine the extent to which findings on the phonetics of lexical self-repair generalize to Colombian Spanish. Subsidiarily, by informing on the specific research questions listed in 1.4, with reference to temporal organization particularly, I contribute to the ongoing discussion regarding the extent to which repair semantics, as classified above, can inform our understanding of the temporal dynamics of repair.

By reporting on the effects of reparandum completeness and the duration distributions of both Target-to-cut-ff and Cut-off-to-repair I help update our knowledge on speech monitoring processes in spontaneous repair, specifically by evaluating whether troublesome lexical items detected earlier are phonetically different in their temporal make-up from those detected later. In doing so, I inform the distinction between trouble sources detected during ‘inner speech’ monitoring and those identified during ‘overt speech’ monitoring. Lastly, the specific sub-question on the impact of editing terms in self-repair, contribute to ratifying the relevance of analysing such editing expressions in repair sequences.

An additional contribution of this thesis, and of this chapter specifically, lies in its methodological approach and the statistical decisions undertaken. By implementing a Bayesian framework and integrating machine learning techniques, I contribute to the phonetic sciences by revisiting key aspects of repair through the lens of innovative statistical methodologies, which have gained prominence in other disciplines but have only recently been incorporated into linguistic research.

The remainder of this chapter is structured as follows. In Section 3.2 I provide a detailed discussion of findings from other languages on the temporal aspects and temporal organization of repair. In section 3.3 (Method), I describe the data set (Section 3.3.1); how each repair was segmented (Section 3.3.2); the classification procedures used to create individualized predictors (Sections 3.3.3 and 3.3.4); and the specifics of the statistical models that were fitted for the Bayesian regressions and the clustering analysis (Sections 3.3.6 and 3.3.5). In Section 3.4, I present the results and show how the outputs from both statistical approaches provide complementary insights into the dynamics of lexical self-repair and its temporal organization. In Section 3.5, I discuss the implications of these findings, including the potential impact of the proposed models and directions for future research, concluding the chapter.

3.2 Temporal Organization of Repair in Other Languages

In this section I provide an overview of studies that have explored the temporal organization of repair on languages different from Spanish. Although this paper focuses on lexical-self repair, I include research that investigated other types of repairs and speech for two main reasons. The first one is that some of these studies present findings that apply to elicited phonological self-repair (e.g., Nooteboom & Quené, 2017, 2019), or repairs under time-pressure manipulation (e.g., Oomen & Postma, 2001) and so on. It is worth investigating whether such findings generalize to other repair types, such as lexical repairs, and other types of speech, including spontaneous language. Secondly, repair has not been widely studied cross-linguistically, with some of the most influential work in the area focusing on a single language, Dutch[15]. Thus, investigations like this one, which explore the phenomenon in a new language or variety, contribute to building broader, more general findings. By adopting this approach, in the following paragraphs, I present all major contributions on the temporal make-up of repair that informed the specific objectives of this study.

[15] Investigations on repair on Dutch include those by Levelt, 1983; 1989; Levelt & Cutler, 1983; Nooteboom, 2010; Nooteboom & Quené, 2019; Nooteboom, 2005a, 2005b; Nooteboom & Quené, 2017; Oomen & Postma, 2001; Plug, 2011, 2015, 2016; Plug & Carter, 2013, 2014; Quené & Nooteboom, 2024; Tydgat et al. (2012). At a smaller scale, we can find research on repair on German by Seyfeddinipur et al. (2008) and on English by Kapatsinski (2010), Blackmer & Mitton (1991), Schegloff et al. (1977) and Goffman (1981).

3.2.1 Repair Timing and Speech Monitoring

Repairs have served as a good basis for studying processes of self-monitoring of speech through features related to the phonetic behaviour of both the reparandum and the repair (e.g., Blackmer & Mitton, 1991; Hartsuiker & Kolk, 2001; Levelt, 1983; Nooteboom, 2010; Nooteboom, 2005a; Nooteboom & Quené, 2017, 2019; Oomen & Postma, 2001; Plug, 2016; Plug & Carter, 2014; Quené & Nooteboom, 2024). Regarding measurements of time and temporal organization, one of the main questions that researchers have addressed is whether speakers interrupt, and repair errors differently based on when a repair target is identified. Levelt's (1989) and Levelt et al's. (1999) Dual Loop Perceptual Theory of self-monitoring proposes two stages: one in which errors are detected in *inner speech* (pre-articulatory), monitoring the speech plan, and one another that identifies trouble targets after speech initiation (post-articulatory), during *overt speech* monitoring. Hartsuiker & Kolk's (2001) contributed to strengthening the Dual Loop Theory and formalized it as a computational model. They simulated empirical data on the distribution of the Target-to-cut-off and Cut-off-to-repair intervals and the effect of speech rate on these intervals. Based on these simulations, they proposed a time course of planning, interruption, and repair for both segmental-level (phonemes) and higher-order units (syllables, phonological words, and phrases) (Hartsuiker & Kolk, 2001, p. 126).

This distinction between *late* and *early* detection of trouble has important implications for the temporal organization of repair, particularly in light of the MAIN INTERRUPTION RULE (MIR) (Levelt, 1989; Nooteboom, 1980). The MIR's key proposal holds that speech is interrupted by speakers immediately as they become aware of any issue in their speech, implying that any delays in repair are the consequence of late detections of trouble (Levelt, 1989, p. 481).

The first study to evaluate the timing of repairs in spontaneous speech was conducted in English by Blackmer and Mitton (1991), with a specific focus on the intervals between error detection and repair initiation. They examined the relationships between Target-to-cut-off and Cut-off-to-repair[16], while also considering the overall Target-to-repair duration. They tested Levelt's (1989) and Nooteboom's (1980) Main Interruption Rule (MIR) and noted how complex the picture is when it comes to describing spontaneous speech. Their study included 1,525 repairs of different nature, (e.g., covert and overt repairs, conceptually-based repairs, such as appropriateness repairs; production-based repairs, like phonological repairs; and repairs involving editing terms). These were extracted from 438 conversational turns produced by 61 callers to a radio show (Blackmer & Mitton, 1991, p. 184).

Results indicate that both intervals had a wide distribution; however, some important patterns were found regarding the nature of the repairs that were analysed. Many of their cut-offs came very fast (i.e., short Target-to-cut-off intervals) and the same happened with their resumptions, with repairs often happening before 100 ms (Blackmer & Mitton, 1991, p. 186). Interestingly, 12.4% of their repairs had Cut-off-to-repairs durations at 0 ms, which suggests that speakers sometimes plan the repairs before they interruption occurs (Blackmer & Mitton, 1991, p. 185). They concluded that production-based errors were repaired faster than conceptually-based ones (errors in meaning or appropriateness issues) with the former averaging around 548 ms and the latter 890 ms. When observing these findings alongside the overall Target-to-repair interval, Blackmer and Mitton propose (1991, pp. 190-193) that repair takes longer for

[16] Error-to-cut-off and error-to-repair in their study.

conceptual and appropriateness repairs, likely due to the additional cognitive effort required for processing such problematic constructions. The latter also challenges the MIR (Levelt, 1989; Nooteboom, 1980) in that the occurrence of immediate repairs (i.e., 0 ms Cut-off-to-repair intervals) does not hold in Levelt's (1989, pp. 473-474) timing estimates. If such calculation were correct, there would be 'no time available to plan a repair before the cut-off, and a repair cannot be ready at the time of interruption' (Blackmer & Mitton, 1991, p. 190); yet this is exactly what happened with some of their resumptions.

Building on the idea of two stages of speech monitoring and their observed implications, Nooteboom (2010)[17] and Plug and Carter (2014) studied spontaneous and elicited phonological repairs on Dutch[18] to empirical support to the dual loop theory (Hartsuiker & Kolk, 2001; Levelt, 1989; Levelt et al., 1999), among other objectives. Nooteboom (2010) studied both experimentally elicited spoonerisms (e.g., mispronunciations such as *barn door* as *darn bore*) elicited by means of the SLIP technique (Spoonerisms of Laboratory-Induced Predisposition) (Baars & Motley, 1974) and spontaneous repairs extracted from the Utrecht Corpus (See Nooteboom, 2010, pp. 219–220 for details).

[17] This paper also presented results on editing terms and prosodic marking in repair. I report the implications of such findings in Chapter 4, which is dedicated to the analysis of prosodic marking, and Section 3.2.2 of the present chapter, in which I discuss the literature related to editing terms.

[18] Speech error repairs in Nooteboom's and Nooteboom and Quené's several papers.

Plug and Carter (2014) focused on spontaneous repairs, with a sample of 368 instances extracted from the Spoken Dutch Corpus[19] (Oostdijk, 2002). For these repair sequences, they measured how the articulation rate of the repair compares to that of the reparandum, while simultaneously evaluating the effects of timing and lexical frequency.

Together, the findings by Nooteboom (2010) and Plug and Carter (2014) provide evidence for both *early* and *late* detection of trouble in speech production, supporting key predictions of the Perceptual Loop Theory. These results reinforce the view that self-monitoring operates at multiple stages, both before and after articulation.

Nooteboom (2010) reported that speakers tend to detect and correct errors in inner speech before they surface, resulting in faster and less noticeable repairs. In contrast, when errors are detected in overt speech, repairs take longer and are more likely to be accompanied by editing expressions. This supports one of Nooteboom's most relevant claims, namely that inner speech monitoring attempts 'to prevent errors in inner speech from becoming public' (Nooteboom, 2010, p. 216); inner monitoring operates under a sort of time pressure, which is not the case for overt speech because at that stage, the problem is already evident, so there is no time pressure to try to prevent it.

[19] The Spoken Dutch Corpus (*Corpus Gesproken Nederlands*; CGN) is available to academic researchers through the European Language Resources Association (ELRA). Further information can be found at: https://www.elra.info/en/catalogues/language-resources/catalogue-of-lrs/?id=494

Based on Hartsuiker et al.'s (2005) idea of attention control during monitoring, Nooteboom (2010) discussed that in spontaneous speech there are more repaired completed errors than interrupted (incomplete) errors, whereas in SLIP repairs there are more interrupted errors than repaired completed errors. This pattern suggests that attentional engagement during overt speech monitoring in spontaneous language is not necessarily lower than during inner speech monitoring; rather, both stages of monitoring appear equally engaged, but are modulated by the objective of the interaction and the resulting type of speech. Specifically, inner monitoring may be more active in experimental settings, while overt monitoring appears to be more crucial in spontaneous speech. Nooteboom argues that this pattern is possible because, in elicited errors, speakers do not need to monitor for lexical conflicts, whereas monitoring for lexical appropriateness is a key aspect of spontaneous communication (Nooteboom, 2010, pp. 231-232).

Plug and Carter (2014) findings align with those of Nooteboom (2010). They reported that the timing of the repair, whether it was caught early or late, had a significant impact on the Cut-off-to-repair interval. Not only did they find that early repairs had shorter Cut-off-to-repair durations, but they also found that these repairs were articulated at a faster rate (Plug & Carter, 2014, p. 61). This supports Nooteboom's (2010) idea that early repairs are executed under time pressure, unlike errors that have already become public. In addition, Plug and Carter (2014) found that higher-frequency words were repaired more quickly than lower-frequency ones in their examination of lexical frequency. They argue that this could be a consequence of the automaticity associated with frequently used words (See also Kapatsinski, 2010).

Nooteboom and Quené (2017, 2019) and Quené and Nooteboom (2024) have investigated the timing and temporal organization of repair, weighting various factors influencing their temporal profiles. Their main purpose has been to advance understanding of speech monitoring and mental speech preparation through elicited phonological repairs in Dutch. Across their different studies, they have focused on Target-to-cut-off[20], Target-to-repair[21] and Cut-off-to-repair durations, either conjointly (Nooteboom & Quené, 2017, 2019) or by centring on a single interval (e.g., Quené & Nooteboom, 2024). They have studied the role of auditory feedback during post articulatory error detection, with findings suggesting that detection does not depend on audition (Nooteboom & Quené, 2017, p. 19). They also examined selective attention based on word structure (with errors occurring word-initially and word-medially) and reported that repairing takes longer for medial than for initial errors and that detection rate decreases from early to late within word forms (Nooteboom & Quené, 2019, p. 43). More recently they focused on mental speech preparation by revisiting some of their previous findings and data in the light of new analyses (Quené & Nooteboom, 2024). In the following paragraphs I present these studies in more detail to describe these findings have informed this paper's specific objectives.

Quené and Nooteboom (2024) re-analysed pooled data extracted from six different data sets collected by the authors since 2005 using the SLIP technique, obtaining a final collection of 1,803 repairs (see (Nooteboom & Quené, 2008) and for details on the experiments, and Quené and Nooteboom (2024) for a full account of the selection of the six data sets used for the re-analysis). This study focused on the distribution of the Target-to-cut-off interval with results corroborating earlier findings by Nooteboom and Quené (2019) that Target-to-cut-off has a bimodal distribution, with two peaks of dispersion separated by approximately 460 ms. This gap

---

[20] Error-to-cutoff in their studies.

[21] Error-to-repair in their studies, which is different from the Cut-off-to-repair because this Error-to-repair is the sum of the Error-to-cut-off-and the Cut-off-to-repair.

is assumed to reflect the delay between the two stages of self-monitoring (Quené & Nooteboom, 2024, p. 9). From Nooteboom and Quené (2019, p. 54), we can also estimate the Target-to-cut-off duration, with it running from 0 ms to more than 1,000 ms. These results, showing wide distributions and bimodality, are again consistent with the proposal by Levelt (1989), Levelt et al (1999) and Hartsuiker and Kolk (2001) that early repairs follow error detection during 'inner speech' monitoring, while late repairs follow error detection during 'overt speech' monitoring. The distinction impacts Cut-off-to-repair duration as well, such that repairs following late error detection in overt speech monitoring require more planning time (Nooteboom & Quené, 2019). In fact, results on the Cut-off-to-repair distribution confirmed that this interval is also wide, with durations starting at 0 ms[22], as also found by Blackmer and Mitton (1991) in English, and reaching up to around 1,000 ms (Nooteboom & Quené, 2019, p. 53). In this line of thought, and as acknowledged by the authors, although speakers can interrupt and repair very quickly, delaying the cut-off and/or the initiation of repair is also an option. This seems likely for a number of reasons, including according to Nooteboom and Quené (2019, p. 54), the non-availability of an immediately accessible repair. Together with the above-mentioned findings, there is strong empirical support for the idea of strategic delay in interruption (e.g., Blackmer & Mitton, 1991; Seyfeddinipur et al., 2008; Tydgat et al., 2012). I will now review relevant studies that have explored the possibility of delaying repairs, as well as those that have documented cases of immediate repair following interruption.

[22] Nooteboom and Quené (2019, p. 53) explain that these Cut-off-to-repair times might be 'censored at zero' meaning that hide quite a few cases where 'the actual moment a repair came available to the mind of the speaker fell a varying amount of time before the moment of interruption'. In these cases, if the speaker planned the repair before interrupting, the real duration might have been negative, but it gets recorded as zero.

Seyfeddinipur et al. (2008) studied the Cut-off-to-repair intervals of 510 overt repairs from a corpus of German speech disfluencies that included video recordings of twelve participants. They explain that speakers, at least on some occasions, decide to privilege fluency over accuracy; the decision is made on a moment-to-moment basis, depending on the impact the troubled speech will have on the flow of conversation (2008, p. 841). The findings support the proposed hypothesis of Delayed Interruption for Planning (DIP), which holds that 'when speakers detect trouble, they do not interrupt but start replanning while continuing to speak according to their original plan' (Seyfeddinipur et al., 2008, p. 838; cf. Levelt's (1989) and Nooteboom's (1980) MIR hypothesis, which suggests immediate interruption upon detection of trouble). If this is the case, and Seyfeddinipur et al. (2008) are correct, repair sequences with long Target-to-cut-off but short Cut-off-to-repair intervals would be expected to emerge in the Colombian Spanish data as well.

Also, in line with the idea of delaying repair, Tydgat et al. (2012) reported lingering in the initiation of repair given the relationship between reparandum and repair in experimentally elicited repairs in Dutch. They used a picture naming task in which images were sometimes suddenly replaced, with the aim of evaluating whether initial lexical and/or phonological activations hindered or facilitated repair (Tydgat et al., 2012, p. 218). The experiments resulted in reparanda being either skipped, interrupted, or completed. They took this information along with measured picture naming latencies to report that in the case of semantic relatedness, abandoned (incomplete) and skipped reparanda (both equivalent to an 'early' detection of trouble) facilitated repair in comparison to completed items. For lexical items that were related in their phonological form, only skipped items eased repair, meaning that the articulation of even a few segments of a phonologically similar lexical item hindered (delayed) repair (Tydgat et al., 2012, p. 225). They concluded that semantic relatedness between reparandum and repair facilitated production while phonological relatedness between the two interfered with it (Tydgat et al., 2012, p. 228). If this is correct, at least for semantically related items, incomplete reparanda with short Target-to-cut-off would be expected to be repaired quickly, and cases of delayed initiation of repair can also emerge because of the absence of a suitable candidate, as was the case with their phonologically similar lexical items, which interfered with the production and, consequently, delayed the repair.

As explained earlier, in their bimodal distributions, Nooteboom and Quené (2019) found early interruptions but also Target-to-cut-off intervals that were far above 1,000 ms. In line with Seyfeddinipur et al.'s (2008) reasoning, they have pointed out that this delaying strategy could be a feature of spontaneous speech (Nooteboom & Quené, 2019, p. 54). As highlighted by Nooteboom and Quené (2017, p. 20), there is no *a priori* reason to expect that their findings on the timing of phonological error repairs will generalize to lexical repairs, as different processing mechanisms are involved, but the question is still worth investigating. Also, as the authors have acknowledged, elicited data is 'appallingly low' (Quené & Nooteboom, 2024, p. 1); and so are spontaneous collections of repairs in other languages. It is still highly relevant to explore whether the patterns found in experimentally elicited repairs hold in new spontaneous data sets and across different repair subtypes. We have already seen that in spontaneous phonological repairs (Plug & Carter, 2014) early and late detection were correlated with fast and slow repairs, respectively. Blackmer and Mitton (1991) also found a wide distribution of their Target-to-cut-off intervals in their conversational repairs, so they found no evidence to reject the idea an inner monitoring system. According to Levelt's (1989) and Levelt et al.'s (1999) Dual Loop Perceptual Theory of self-monitoring, the two stages of error detection proposed are to happen in both elicited and spontaneous language since they are not thought to be exclusive to a certain type of speech. Given the empirical support found, the expectation is that instances of early and late detection of trouble will emerge in our spontaneous data as well. In fact, Quené and Nooteboom (2024, p. 4) acknowledge that the 'early' and 'late' split of approximately 460 ms in their repairs, might well be an underestimation, consequence of the artificial limitations of elicited errors. Thus, the actual delay between 'early' and 'late' identification in normal communication conditions could in fact be greater those 460 ms. Still, Plug and Carter (2014) did not find evidence of bimodality in their Target-to-cut-off intervals in spontaneous phonological repairs, which could be a consequence of these being repairs similar to those classified Blackmer and Mitton (1991) as 'production-based' repairs, which according to them are usually delt with faster than conceptually-based and appropriateness repairs. As noted by Nooteboom (2010, p. 31) monitoring for lexical errors is a key aspect of spontaneous communication. In elicited phonological repairs, speakers neither need to monitor for lexicality of the erroneous items, nor have the time to do so. However, it seems that even in spontaneous phonological repairs, it is plausible that these errors are repaired earlier and fast. Thus, spontaneous lexical repairs can offer meaningful insights into this debate by informing whether what happens in Colombian Spanish mirrors Seyfeddinipur et al.'s (2008) and Blackmer and Mitton's (1991) findings on German and English.

Taken together, if Nooteboom's (2010) and Quené and Nooteboom's (2024) predictions for spontaneous lexical repairs are right, long Target-to-cut-ff intervals will emerge in this thesis' unscripted lexical repairs, and the gap between early and late detection could be even wider than reported for their elicited phonological repairs.

In this chapter, I hope to add valuable insights to the discussion by reporting on the temporal frame of both Target-to-cut-off and Cut-off-to-repair intervals in conversational speech, in relation to the 'early' and 'late' identification debate in speech monitoring, through an analysis of lexical self-repair.

3.2.2 Repair Semantics and Editing Terms in self-repair

As we have already discussed, many of the investigations on the temporal organization of repair have focused on informing processes of speech monitoring and defining how the moment of detection of trouble influenced the timing of the repair. In addition to describing the temporal make-up of repair, some authors have sought to deepen our understanding on the reasons that underlie a specific behaviour of speakers when timing the interruption and resumption of their repairs (e.g., Blackmer & Mitton, 1991; Nooteboom, 2010; Oomen & Postma, 2001, 2002; Plug & Carter, 2014; Seyfeddinipur et al., 2008; Tydgat et al., 2012). Similar to the studies mentioned above, the semantics of the repair is one of the linguistic features that has been studied as a factor influencing the temporal organization of repair.

The presence of editing expressions has also been observed by some authors, but more as a functional part of repair (i.e., Levelt, 1989) rather than as a factor itself. In this section I will discuss the findings that motivated the selection of the different repair subtypes and the use of editing terms, as possible factors influencing the temporal make-up of lexical self-repair.

Levelt (1983), Levelt and Cutler (1983), and Plug (2016) evaluated the effects of semantics in self-repair. Levelt (1983) and Plug (2016) assessed such effects on the timing and temporal organization of repair sequences, while Levelt and Cutler (1983) focused on prosodic marking in repair. Levelt (1983) and Levelt and Cutler (1983) split their repair collections between issues of *appropriateness* and *errors*. As explained earlier, the former refers to the resolution of problematic constructions, making them more appropriate for the context, while the latter correct actual mistakes. Levelt (1983) found that error repairs were more likely to be interrupted given that they carried more relevant information for the listener than appropriateness repairs. He proposed that the trouble source in appropriateness repairs might require elaboration (in qualitative terms) rather than replacement (Levelt, 1983, p. 64), and for that reason, these repairs are less likely to be interrupted.

Regarding prosodic marking, Levelt and Cutler (1983) also found an effect of semantics in their repairs. They reported that error repairs are more likely than appropriateness ones to be ‘prosodically marked’ in that repairs made after an error are produced with higher pitch and intensity. Following these findings, Plug (2016) revisited the effects of semantics by investigating the distinction between appropriateness and factual errors, but added to the discussion by further dividing errors into factual and linguistic repairs. As described earlier in this chapter, errors can be linguistically incorrect or factually wrong. The purpose of including such sub-categories derives from the question as to whether distinct levels of processing involved in error detection could lead to different temporal organizations (Plug, 2016, p. 522) (see also Plug & Carter, 2013). Plug’s (2016) analyses were extensive in that he investigated the relationship between semantics, reparandum completeness and lexical frequency. He found that high-frequency reparanda are more commonly completed than their low-frequency counterparts. Regarding semantics, Plug (2016) reported that factual errors and appropriateness repairs ‘show little difference in temporal organization’ (Plug, 2016, p. 539). Linguistic errors tend to involve fewer incomplete reparanda, with those reparanda featuring items of higher lexical frequency. Levelt (1983) found that errors were more often incomplete than appropriateness repairs, but he did not separate linguistic from factual errors; Plug (2016, p. 535) observes that it is possible that the high frequency associated with grammatical words (i.e., prepositions, particles, pronouns), usually repaired in linguistic errors, could have played a role in the timing of these repairs and might explain the discrepancy between his findings and Levelt's (1983), which did not distinguish between linguistic and factual repairs. These contrastive findings highlight the need for further studies exploring the differences between error and appropriateness repairs, as well as to differentiating linguistic errors from other two repair subtypes. On the 'early' and 'late' detection contrast, Plug (2016) reported that spontaneous repairs in which the target word is completed before interruption have lower Cut-off-to-repair durations with a tendency for repairs that are started early to be completed faster, and for repairs that are initiated late to be completed more slowly, similarly to what was reported earlier by Plug and Carter (2014) on spontaneous phonological self-repair (Plug, 2016, p. 538). Overall, these findings are an invitation to continue exploring these timing relationships in spontaneous lexical repair in new data sets as well to investigate whether similar patterns for linguistic and factual errors hold in new data sets.

Repairs with editing terms, such as *bueno* 'well' in repair sequences such as *cuatro, bueno, tres*, 'four, well, three'; may complicate the analysis of the temporal organization of lexical self-repair. However, in repair, editing terms play a key role in signalling to interlocutors that there is trouble (Levelt, 1989, p. 482) and there is evidence proving that different subtypes of repair are associated with distinct editing term choices (e.g., James, 1972; Levelt, 1983). There is an agreement that these expressions are a functional and important part of repair sequences (e.g., Levelt, 1989; Nooteboom, 2010; Nooteboom, 2005b; Plug, 2016). With respect to the occurrence of such expressions, it seems reasonable to expect that Cut-off-to-repair intervals containing an editing term are on average longer than silent ones. However, if we keep in mind, as noted by Plug (2016), the complex relationship between the various measures of informativeness in repair, there might be more to it than a simple consequential delay in the repair because the articulation of such expressions consumes time in between the interruption and the repair. In support of this, we can take Levelt's (1983) semantic analysis of self-repair where he reported that errors were more often followed by editing expressions than appropriateness repairs. For phonological repairs, Nooteboom (2010), whose focus was on the role these expressions played in overt speech monitoring, informed that repairs of completed errors take longer and are more likely to be accompanied by editing expressions. This aligns with his idea that speakers will attempt to direct the attention of their audience to the speech errors they have detected in their overt speech; if an error is already public (i.e., fully articulated), it does not have any additional consequence if the exchange becomes longer and it is accompanied by editing expressions.

In spontaneous lexical self-repair, Blackmer and Mitton (1991) also commented on the role of editing expressions in their English repairs and suggested that 'slower' repairs (i.e., conceptually based) are more likely than 'faster' repairs (i.e., sound-form repairs) to be followed by editing terms. Seyfeddinipur et al. (2008, p. 841) suggested for their German conversational data that speakers can use editing terms strategically when they cannot repair quickly, and end up running out of time; by doing so, they also minimize misunderstandings, since they inform the listener that they are experiencing difficulties in their speech, a possibility that had also been proposed by Levelt (1989).

3.2.3 This Study

To further our understanding of the temporal organization of lexical self-repair, in this study I explored the distributions of Target-to-cut-off and Cut-off-to-repair intervals in a sample of self-repairs extracted from unscripted conversations in Colombian Spanish. In particular, the study addresses the following general hypotheses:

- **HYPOTHESIS A:** Complete reparanda will result in longer Target-to-cut-off and Cut-off-to-repair durations than incomplete reparanda. Incomplete reparanda are associated with early detection of trouble and interruption, while complete reparanda correspond to late detection of trouble and interruption.
- **HYPOTHESIS B:** A semantic classification of repairs, distinguishing between linguistic and factual error repairs versus appropriateness repairs, will predict Target-to-cut-off and Cut-off-to-repair durations.
- **HYPOTHESIS C:** The presence of editing terms will result in longer Target-to-cut-off and Cut-off-to-repair durations and editing terms will more frequently accompany appropriateness repairs, factual error repairs and repairs with complete reparanda.
- **HYPOTHESIS D:** The Target-to-cut-off and Cut-off-to-repair distributions will include durations close to 0 ms as well as durations around 1,000 ms. Individual inspection of each interval is expected to show bimodality, and clustering analysis will reveal multimodality across the two intervals.

HYPOTHESIS A is consistent with Nooteboom's (2010), Plug & Carter's (2014) and Plug (2016) observations on the temporal organization of phonological and lexical repairs. It also supports Levelt's (1989) and Levelt et al.'s (1999) Dual Loop Perceptual theory and Hartsuiker and Kolk's (2001) Perceptual Loop Model, which propose that there are two stages of speech monitoring, and delayed repairs are the consequence of late detection. HYPOTHESIS B is in line with work by Levelt (1983), Plug and Carter (2014) and Plug (2016) who found effects of the semantically motivated subclassification of repairs. It is also consistent with Blackmer and

Mitton (1991), who classified appropriateness and conceptually based repairs as having a different timing compared to production-based repairs, which are more similar to the linguistic repairs described by Plug (2016).

HYPOTHESIS C is consistent with Blackmer and Mitton (1991), who reported that conceptually based repairs were far more likely to be accompanied by editing terms than production-based repairs. Based on Blackmer & Mitton's (1991) idea of reformulation (i.e., speech halt and empty buffer) in slower repairs (e.g., appropriateness repairs) and Plug's (2016) suggestion of automaticity associated with linguistic errors, both factual error repairs and appropriateness repairs are more likely to require such reformulation than linguistic error repairs, and therefore to be accompanied by editing terms. This view contrast with Levelt's (1983) findings that errors were more frequently followed by editing expressions than appropriateness repairs.

HYPOTHESIS D aligns with findings by Nooteboom (2010), Nooteboom & Quené (2017, 2019) and Quené and Nooteboom (2024). It follows their proposal of both Target-to-cut-off and Cut-off-to-repair having wide distributions, with Target-to-cut-off confirmed bimodal. These wide distributions are expected to be reflected in the clustering analysis with configurations including both early-detection and early-resumption and late-identification combined with late-correction. These two strategies are consistent with the argument related to two temporal organizations, explained in terms of *inner* versus *overt* speech detection (e.g., Nooteboom, 2010). Moreover, within these wide distributions it is feasible to have Cut-off-to-repair durations starting at 0 ms (Blackmer & Mitton, 1991; Nooteboom & Quené, 2019) and up to a 1,000 ms (Nooteboom & Quené, 2019). This hypothesis is further consistent with the possibility of postponing interruption if a repair is not available, supporting the idea of strategic postponement of repair, as reported by Seyfeddinipur et al. (2008), who observed a delay favouring fluency, and with Tydgat et al. (2012), who found that semantic relatedness facilitated repair, while phonological relatedness hindered it and delayed it.

With the aim of fulfilling this study objectives and confirming or rejecting the above-listed hypotheses, this chapter examines indications of bimodality following Nooteboom and Quené (2017, 2019) and Quené and Nooteboom (2024). The search for bimodality together with the analysis of the moment of interruption and completeness of the reparandum in relation to the timing of the repair also informs the frequently observed 'early' and 'late' stages of speech monitoring (Hartsuiker & Kolk, 2001; Levelt, 1983).

This research additionally considers the relationship between Target-to-cut-off and Cut-off-to-repair, taking into account the existence of multiple semantic sub-types of repairs, including, as suggested by Plug (2016, p. 539), the distinction between factual and linguistic errors. Finally, the study investigates evidence of cut-off delaying strategies as proposed by Seyfeddinipur et al. (2008) and Tydgat et al. (2012), while weighing the presence versus absence of editing terms based on findings by Levelt (1983), Blackmer and Mitton (1991) and Plug (2016).

3.3 Method

This section is devoted to presenting all relevant decisions that were made regarding the methodology adopted for the relevant analyses and the procedure that was followed. First, I describe in detail the processes for data selection, segmentation and classification (Sections 3.3.1 to 3.3.4). Then, in Section 3.3.5 I introduce the variables that were incorporated into the statistical analysis, including the relevant choices used for prior class for each parameter that was included in the Bayesian models. Finally, Section 3.3.6. defines the concept of clustering in machine learning and presents the parameters that were set for performing the clustering.

3.3.1 Timing Analysis Data Set

Data for this report include 404 instances of self-initiated lexical self-repair extracted from sub-corpora of the UCSI Corpus. The corpus is composed of unscripted speech from the different dialects and subdialects of Colombian Spanish, based on the dialectal classification by Ruiz Vásquez (2020). Conversational speech was gathered through two interactive tasks (See sections 2.3.1 and 2.3.2 for reference). Repairs were sampled from recordings of both tasks completed by 36 speakers, from 5 dialects of Colombian Spanish. For this thesis, 431 instances of lexical self-repair were processed. Before running the analyses, a few repairs were excluded from the sample due to difficulties in their semantic classification, which I explain in Section 3.3.3, where I present the detailed procedure for classifying each item as a linguistic, factual or appropriateness repair. In addition, while processing the repairs, I identified two types of repair sequences that impacted the duration of at least one of the intervals of interest. First, there were instances in which speakers attempted to repair the reparandum more than once; such cases were coded as having *Several Attempts* at repair (coded as SA). Example (13) illustrates such behaviour. In addition, sometimes speakers addressed the issue occurring with their speech quite prominently, and made comments highlighting it, which in some cases led to exceptionally long Cut-off-to-repair intervals. These cases were coded as speakers giving *Attention to Error* (coded as AE); 83(14) illustrates one of these cases. Either of these ways of interacting around the errors or infelicities (n = 27) were also excluded from the final data set. After removing the above-mentioned repair sequences from the sample, the final data set for timing analysis held 405 repairs.

(18) Self-initiated, self-repair including SEVERAL ATTEMPTS at repair

Spanish (Vera Diettes sp_023_024_conv_kv_part6)

01	A:	Pero como **tú decías;** o, como, como, **hemos hablado.**
		*But as **you said**, or, as, as **we have talked about**.*

(19) Self-initiated, self-repair with ATTENTION TO ERROR

Spanish (Vera Diettes sp_023_024_conv_kv_part6)

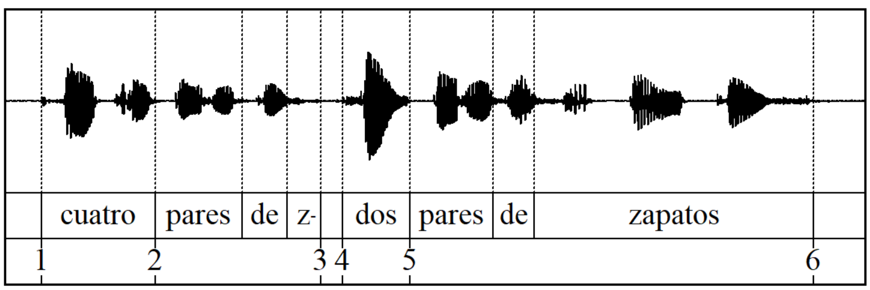
01	A: →	Que apoyo lo de la **pena de muerte,** y si- ¡Dizque la pena de muerte! Yo ya estoy muy violenta.
		*That I support the thing about the **death penalty,** and if- Death Penalty, really? I'm way too aggressive.*
02	B:	Yo también.
		Me too.
03	A:	Lo de la **cadena perpetua.**
		*The thing about **life imprisonment**.*

3.3.2 Segmentation

Each repair was orthographically transcribed by me, the author, a native speaker of Colombian Spanish, and then semi-automatically transcribed in Praat (Boersma & Weenink, 2021) using the EasyAlign Spanish extension (Goldman & Schwab, 2014). Resulting phoneme, word, syllable, and phone tiers were hand-corrected by the author, as illustrated earlier in Figure 6. For the analysis of temporal organization of repair, a point tier was added to locate crucial temporal landmarks. To illustrate this, in Figure 8 we can observe an example in which the target of the repair, *cuatro* 'four', is corrected to *dos* 'two'. The speech is cut off by the speaker at the first sound (i.e., [s]) of the word *zapatos* 'shoes'. The interval between 1 and 3 is the Target-to-cut-off interval; the interval between 3 and 4 is the Cut-off-to-repair interval. All Target-to-cut-off and Cut-off-to-repair durations were extracted for analysis.

Figure 8

Segmented waveform of the lexical factual repair 'Dos pares de zapatos' (Two pairs of shoes).



Note. '1' and '2' delimit the target word; '1' and '3' delimit the Target-to-cut-off; '3' is the cut-off; '3' and '4' delimit the Cut-off-to-repair; '4' and '5' delimit the repaired target word; and '4'and '6' delimit the complete repaired phrase.

3.3.3 Semantic Classification

Each instance was classified as an appropriateness or error repair, using the criteria described in Plug (2016). Repairs with factual inaccuracies, such as *tres, bueno, cuatro* (in English, 'three, well, four'), or linguistic ill-formedness, like [23]￼, *un cartel pequeñito* (in English, 'a sign-y, a small sign'), were classified as error repairs. For all other cases, it was assumed that the first lexical choice was rejected for pragmatic felicity, or appropriateness reasons, like in the repair *una droguería* [24] 'a drug store', instead of *una farmacia* 'a pharmacy'. For an initial subset of 104 instances, the classification was performed by me and an independent second rater, a Colombian Spanish discourse analyst. Classifications matched for 92 repairs (i.e., 90% of the cases), and consensus was reached for 9 out of the remaining 10; one repair was excluded, as the raters agreed it could not be reliably classified. Given these results, for the remaining 301 repair sequences, I made an initial classification, and only doubtful instances were presented to the second rater for consensus.

In total, the 404 items were first classified as errors or appropriateness repairs; subsequently, errors were subdivided into factual and linguistic repairs. Table 7, below, shows a summary of the final semantic classification for the analyses performed.

---

[23] Adding the -er- segment in *cartelerito* is ungrammatical because it does not follow the standard diminutive formation rules in Spanish. The correct morphological form of the diminutive for cartel is *cartelito*, which adheres to the established pattern of attaching the diminutive suffix -ito/-ita directly to the root. This makes *cartelerito* a linguistic error in morphology. While English lacks a formal diminutive system, a comparable error might involve the addition of a suffix, such as 'sign-y'.

[24] Although drug store and pharmacy have the same meaning in Spanish, pharmacy can be perceived as a more refined and, therefore, less frequent alternative in Colombian Spanish. For that reason, it is possible that the speaker in this example repaired to choose the less sophisticated word to better fit the context.

Table 7

Final semantic classification for the analysis of temporal organization.

Repair Type		**Error Type**	
Appropriateness	159		
Errors	246	*Linguistic*	64
		Factual	181
Total	**404**		

3.3.4 Coding for Editing Terms and Completeness

Classifying items according to the presence or absence of editing terms was done following Levelt (1983). Examples with editing terms include words like *o sea* 'I mean' or *bueno* 'well' and filled pauses such as *eee* [e:] and *mmm* [m:], in between the reparandum and the repair. Out of the 404 repair sequences, 118 have editing terms meaning that the great majority of them (i.e., 286) did not include such editing expressions. For *Completeness*, the items in which the first attempt at the target word is stopped before its completion were classified as INCOMPLETE, while items in which the target is pronounced in full were coded as COMPLETE. It is important to keep in mind that the moment of interruption is different from the classification of a target as complete or incomplete. Identifying the moment of interruption helps us calculate the Target-to-cut-off, simultaneously allowing us to calculate how long it takes speakers to interrupt the flow of speech after starting the articulation of the target or reparandum. This interruption can happen immediately, within the same target word; or, it can be delayed, so the interruption of the speech flow takes place at the next phrase following the reparandum, or even later. For instance, recall the example illustrated in Figure 8 (above) *Dos pares de zapatos* 'Two pairs of shoes', where the speaker interrupted the flow of speech at the first sound of the noun *zapatos* 'shoes'. The actual reparandum in that repair sequence is *cuatro* 'four', which was later repaired to *dos* 'two' in the phrase ‘two pairs of shoes’. Therefore, although the word *zapatos* is not articulated in full, as it was prematurely interrupted, the actual target word *cuatro* ‘four’ was classified as a *complete* reparandum because it was fully articulated moments before the actual interruption of the speech flow.

Defining how long it takes speakers to interrupt speech (i.e., calculating the Target-to-cut-off) can be of help for understanding processes of speech monitoring. It is possible speakers have incomplete targets, which are expected to have very short Target-to-cut-off durations in comparison to fully articulated targets (i.e., complete reparanda); however, it is also possible to observe delayed interruptions, as in the case with the repair ‘*Two pairs of shoes*’. This might represent an example of a strategically delayed interruption, which can have an impact on the Cut-off-to-repair duration as well. Consequently, it is important to include the proposed coding (*complete* versus *incomplete* reparanda) as it facilitates clearly distinguishing these cases from instances of delayed interruptions (i.e., even longer Target-to-cut-off intervals).

Following the application of these criteria, the distribution of complete and incomplete reparanda was as follows: 227 instances were classified as having complete reparanda, whereas 177 instances were identified as incomplete.

3.3.5 Statistical Analysis Using Bayesian Modelling in brms

The distributions of the durations of Target-to-cut-off and Cut-off-to-repair were first inspected. Subsequently, analyses were performed in R (R Core Team, 2023) by implementing Bayesian regression models using *brms*[25] (Bürkner, 2017) and the *tidyverse*[26] package (Wickham, 2019).

The Bayesian framework was chosen since it offers a series of advantages in comparison to frequentist models fitted using the *lme4* package (Bates et al., 2015). Firstly, it presents more flexibility in defining models, allowing for complex random effects structures to be fitted without problems (Nicenboim & Vasishth, 2016), which, as also explained by Winter et al. (2023), makes Bayesian models more likely to converge than their frequentist counterparts. Other important advantages of the Bayesian framework are that it offers the possibility of incorporating prior knowledge into the analysis which at the same time allows to quantify the uncertainty about the size of an effect (Vasishth et al., 2018). Table 8 shows the variables entered in the analysis.

---

[25] *brms* is an R package for fitting Bayesian multilevel models using Stan and formula syntax. See: https://paul-buerkner.github.io/brms/.

[26] *tidyverse* is a collection of R packages for data manipulation and visualization, including `ggplot2`, `dplyr`, and others. See: https://www.tidyverse.org.

Table 8

Variables entered into the analysis.

Dependent variables	**Target-to-cut-off, Cut-off-to-repair**
Independent variables	**Completeness** 1. *Complete* 2. *Incomplete*
	Repair Type 1. *Appropriateness infelicities* 2. *Factual errors* 3. *Linguistic errors*
	Editing Terms 1. *Yes* 2. *No*
Group level effects	**Speaker**

Once variables were defined, two main statistical models were fitted to test the hypotheses for the study presented earlier, one for Target-to-cut-off (`mod1`) and one for Cut-off-to-repair (`mod2`). Both models included all predictors and the speaker as the only group level effect (equivalent to the frequentists' random effects). Following Lemoine (2019), priors were set aiming at weakly informative distributions with the aim of regularizing results arising from small sample sizes, as it is the case in many linguistic studies (Winter & Bürkner, 2021, p. 14) and not an exception with phonetic data. The latter also avoids missing one of the major advantages of the Bayesian framework; that is, incorporating prior knowledge into the analysis. By choosing weakly informative or 'regularizing' priors we can incorporate prior knowledge (i.e., quantitative patterns observed in previous research) in a conservative way. The latter also allows us to incorporate 'mild skepticism' into the analyses by introducing regularization, which avoids overfitting without being too skeptical, preventing the model to learn from the data (McElreath, 2020, p. 216). Table 9, below, presents each prior class and the choices made for every parameter.

Table 9

Priors set for the models evaluating the effects of predictor variables on Target-to-cut-off and Cut-off-to-repair.

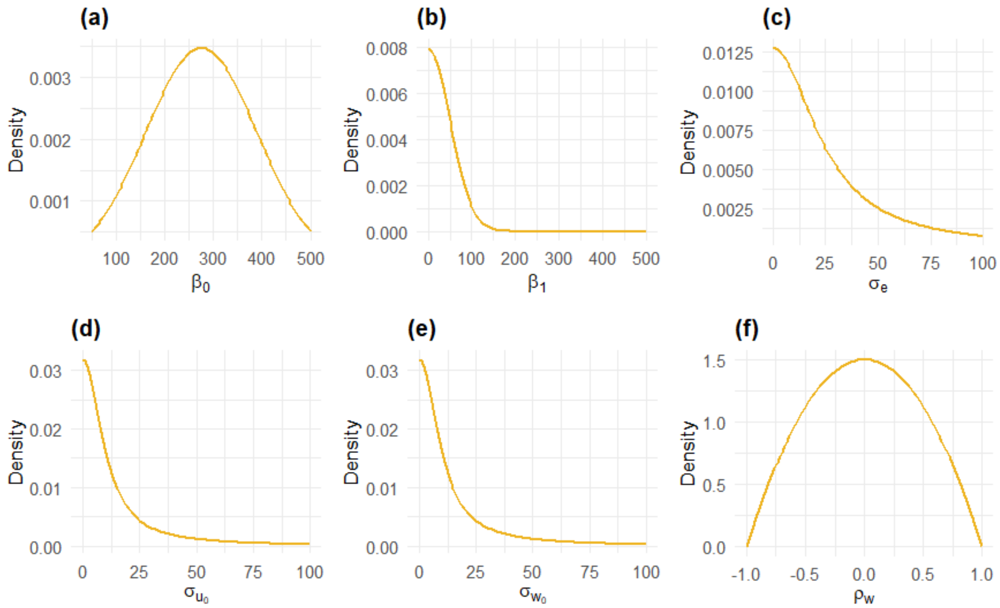
Prior class and settings

`weak_priors <- c(prior(normal(0, 0.5), class = b),`
`prior(normal (5, 0.6), class = Intercept),`
`prior(cauchy(0, 0.25), class = sigma),`
`prior(cauchy(0, 0.01), class = sd),`
`prior(lkj(2), class = cor))`

The specifications given in Table 9 define the different priors for each `class` of parameter. The class `Intercept` (i.e., β_0) are the intercept parameters (dependent variables). The prior for both intercepts were set as *Normal* (275, 115), (*Normal* (5, 0.6), after log transformation) following several reports on these intervals in the context of repair: Hartsuiker and Kolk (2001), Oomen and Postma (2001), Oomen and Postma (2002), Seyfeddinipur et al. (2008), Plug and Carter (2014), Nooteboom and Quené (2017), and Nooteboom and Quené (2019). Although some of these papers reported shorter intervals, I chose this specific prior so that I included the widest range of possibilities reported for both Target-to-cut-off and Cut-off-to-repair, also aiming at standardizing both intervals. The class `b` (i.e., β_1) refers to the slopes in the model; that is, the slopes for the predictors which are the effects that are theoretically of interest (*Completeness*, *Repair Type* and *Editing Terms*). This prior was set at normal distribution centred at zero with SD = 0.50 (*Normal* (0, 0.5); which is a wide prior that allows for an ample range of differences between the conditions. The specifications given in Table 9 define the different priors for each `class` of parameter. The class `Intercept` (i.e., β_0) are the intercept parameters (dependent variables). The prior for both intervals were set as *Normal* (275, 115), (*Normal* (5, 0.6), after log transformation) following several reports on these intervals in the context of repair: Hartsuiker and Kolk (2001), Oomen and Postma (2001), Oomen and Postma (2002), Seyfeddinipur et al. (2008), Plug and Carter (2014), Nooteboom and Quené (2017), and Nooteboom and Quené (2019). Although some of these papers reported shorter intervals, I chose this specific prior so that I included the widest range of possibilities reported for both Target-to-cut-off and Cut-off-to-repair, also aiming at standardizing both intervals. The class `b` (i.e., β_1) refers to the slopes in the model; that is, the slopes for the predictors which are the effects that are theoretically of interest (*Completeness, Repair Type* and *Editing Terms*). This prior was set at normal distribution centred at zero with SD = 0.50 (*Normal* (0, 0.5); which is a wide prior that allows for an ample range of differences between the conditions (Winter & Bürkner, 2021, p. 15). Also, the *Normal* (0, 0.5) prior for all class *b* coefficients allows for a comparison of the effects of each of the variables on the intervals. The standard deviation, class `sigma` (σ_e) can be given a non-informative prior (Lemoine, 2019, p. 918); therefore, I set a *Cauchy* distribution centred at zero with a SD=0,25 (*Cauchy* (0, 0.25). Similarly, the parameter class `sd`, the standard deviation for the group level effects (i.e., σ_u and σ_{wo}), was placed at *Cauchy* (0, 0.01) after Lemoine's (2019, p. 925) descriptions showing the potential of the Cauchy distributions for yielding proper posterior estimates across different sample sizes. Finally, for the correlation parameter (ρ_w), class `corr`, I went for the standard choice LKJ(2) assuming that extreme values are unlike (Vasishth et al., 2018, p. 150). Priors chosen can be visualized in Figure 9.

Figure 9

Prior distributions for the parameters in linear mixed models `mod1` *and* `mod2`.



Note. The prior for the grand mean parameter $\boldsymbol{\beta_0}$ (a) is a normal distribution with mean 275 and standard deviation 115; the prior for the parameter representing the effect of the intercepts $\boldsymbol{\beta_1}$ (b) is *Normal* (0, 0.5); the priors for the standard deviation $\boldsymbol{\sigma_e}$ (c), were set at a *Cauchy* (0, 0.25) distribution and *Cauchy* (0, 0.01) distribution for the standard deviation of the group level effects ($\boldsymbol{\sigma_u}$ and $\boldsymbol{\sigma_{wo}}$) (d and e, respectively); and, the prior for the correlation between the random effects $\boldsymbol{\rho_w}$ (**f**) is LKJ($v = 2$).

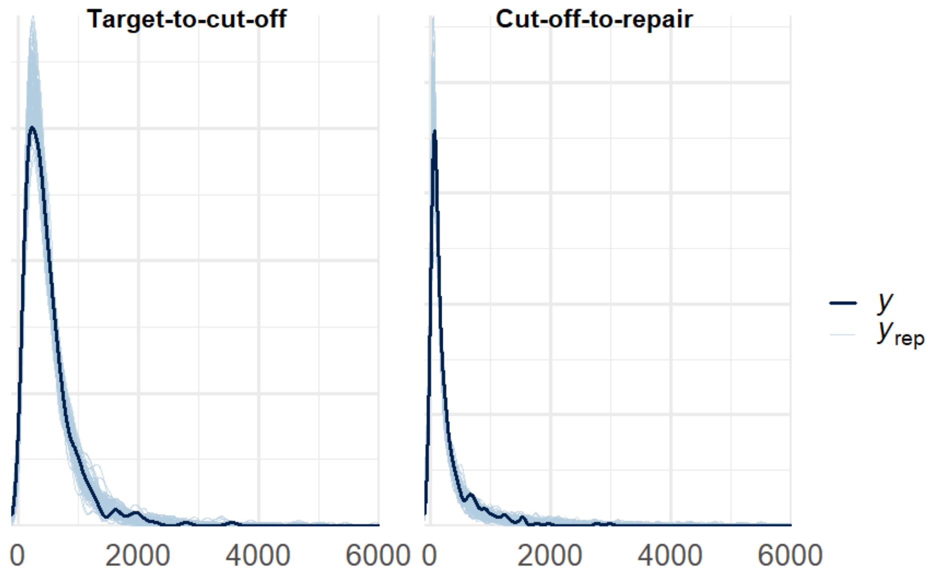
In *brms* the argument `family` relates to the distribution of the response variable. The `family = lognormal()` was used to model for our response variables since it has the best fit for dealing with measurements of time (Bürkner, 2017, p. 8). In addition, the choice of the log-normal distribution was driven by the inherently prone to skewness nature of time-to-event data, where most intervals are shorter, but a few can be substantially longer. The log-normal model effectively captures this skewness, providing a better fit than models assuming a normal or other symmetric distribution. Moreover, sensitivity analyses comparing alternative model families indicated that the log-normal distribution consistently provided a better fit.

The models were estimated via Markov Chain Monte Carlo (MCMC). Four sampling chains were run for 8000 iterations, with a warm-up period of 4000 iterations. Adapt-delta was increased in the `control` argument to avoid a few divergent transitions that emerged during our experimental runs. In the final models, there were no divergent transitions, and all chains mixed well (Rhat = 1.0 for all models). Posterior predictive checks, which assess how well the model matches up with the observed data (Vasishth et al., 2018), show that the predicted and observed data for `mod1` (Target-to-cut-off) and `mod2` (Cut-off-to-repair) have similar distributions (See Figure 3). Therefore, it can be concluded that the models have a reasonable fit. Finally, although the models used log-transformed values, for clarity and ease, results will be presented and interpreted in the original count scale (milliseconds). Back-transformation into milliseconds was done following (1), where e is the base of the natural logarithm (approximately 2.718) and β is the log-transformed coefficient from the model. Full summaries in the models' log-transformed scale can be found in Appendix E.

$$Back - transformed\ coefficient = e^{\beta} \tag{1}$$

Figure 10

Posterior predictive checks for the data. The lines marked y_{rep} refer to the posterior predictive values generated by the models, and the black solid line are the observed data.



3.3.6 Statistical Analysis Using k-means Clustering

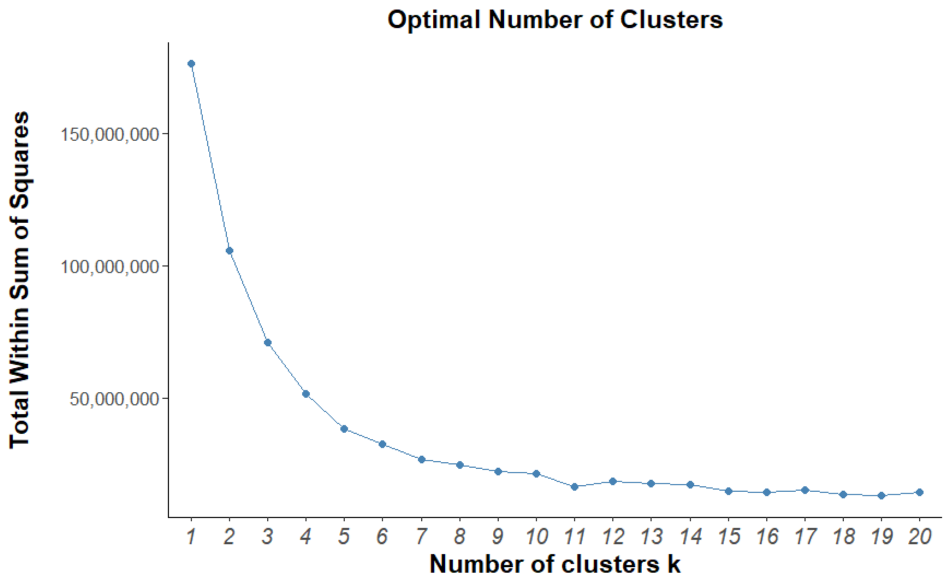
An assessment of the emergent distributions of Target-to-cut-off and Cut-off-to-repair was made using k-means clustering by implementing the R package *stats* (R Core Team, 2024). K-Means one of the most well-studied clustering approaches for finding inherent groupings or clusters in the data based on similarities (Jain, 2010). Unsupervised learning, as that implemented through clustering, involves training algorithms on data without labelled responses (i.e., how many natural clusters or categories are there in the data), allowing the algorithm to identify patterns in the data on its own (Malik & Tuckfield, 2019). One important feature of k-means is that it requires a target number of clusters, or *K*, at the start of the analysis; therefore, choosing the right number of clusters becomes an important part of process. The optimal solution in k-means minimizes the distance between points in a cluster, while maximizes between-cluster variance

(Steffman et al., 2024). The Within-Sum-of-Squares (WSS) method does exactly that by calculating a score that is the sum of the squares of the distances of all points within a cluster (See Malik & Tuckfield, 2019 for details on the calculation of the WSS scores and a comparison between this and other methods). The method assumes that when the number of clusters is much smaller than the optimal number, the total WSS will be much larger. As the number of clusters approaches the optimal value, the WSS decreases significantly (Hardy, 1994), resulting in a sudden drop that creates the 'elbow' point on the graph, hence the name 'elbow method'. We implemented WSS with the package *factoextra* (Kassambara & Mundt, 2020), to identify the ideal value of k (See Figure 11). We chose $k = 3$ as the elbow of the graph, since the value of WSS starts dropping more slowly after $k = 3$. Given that choosing the elbow can be subjective, we also tested other solutions (e.g., $k = 4$ and $k = 2$)[27] but we found the emergent clusters using $k = 3$ as good candidates for the distinction of Target-to-cut-off and Cut-off-to-repair, therefore, $k = 3$ is the clustering solution that we report in this chapter. In consequence, for our analysis, the number of clusters in the optimal solution represents the best category candidates based on cluster distinctions. As a final consideration before running the clustering analysis, we set a random seed, which ensures the reproducibility of our results.

[27] Alternative Bayesian model solutions, as well as additional materials and resources, are available through the OSF repository at https://osf.io/zrqxd/.

Figure 11

WSS versus number of clusters.



3.4 Results

The mean duration of Target-to-cut-off was 497.1 ms (range: 50–3547 ms), and the mean duration of Cut-off-to-repair was 264 ms (range: 1–2983 ms). Based on model intercepts, when all predictors were held constant, the estimated Target-to-cut-off duration was 489 ms, with a 95% Credible Interval (CI) [428, 560 ms]), and the estimated Cut-off-to-repair duration was 82 ms 95% CI [64, 105 ms]). Hartigan's dip test confirmed that the distribution of Target-to-cut-off durations was not significantly multimodal ($D = 0.017$, p = 0.582), supporting the assumption of unimodality. In contrast, for Cut-off-to-repair, the dip test suggested a possible departure from unimodality ($D = 0.027$, $p = 0.039$). A Pearson correlation[28] between Target-to-cut-off and Cut-

[28] See Appendix D for Scatterplot Matrix of variables compared.

off-to-repair durations revealed a small but statistically significant positive association ($r = 0.24$, $p < 0.001$). Although a weak correlation was present, the intervals were treated as independent outcome measures given that they reflect distinct phases of repair processing; namely, interruption timing and repair onset timing.

3.4.1 Relevance of Completeness, Editing Terms and Repair Type on Target-to-cut-off

`Mod1` evaluated the effects of all independent variables on Target-to-cut-off. Results show that both *Completeness* and *Editing Terms* have a significant impact on the duration of this interval, while the different repair types (i.e., appropriateness, linguistic and factual) show relatively small effects. The impact of *Completeness* aligns with expectations, as incomplete repairs are expected to have shorter Target-to-cut-off durations; however, the relationship of editing terms with the timing interval preceding them came as an unexpected and theoretically interesting finding, warranting further examination. In consequence, the relationship found between Editing Terms and Target-to-cut-off was further explored by refining `Mod1` to include interactions between the presence or absence of these editing expressions and the different semantic repair types. This interaction analysis revealed an increase in duration in the Target-to-cut-off when editing terms are present, particularly in the context of appropriateness repairs. In the following paragraphs, I present these results in detail, including the estimated differences, the magnitude of the effects in terms of percentage change, calculated by implementing (2), and the corresponding credible intervals for each variable analysed.

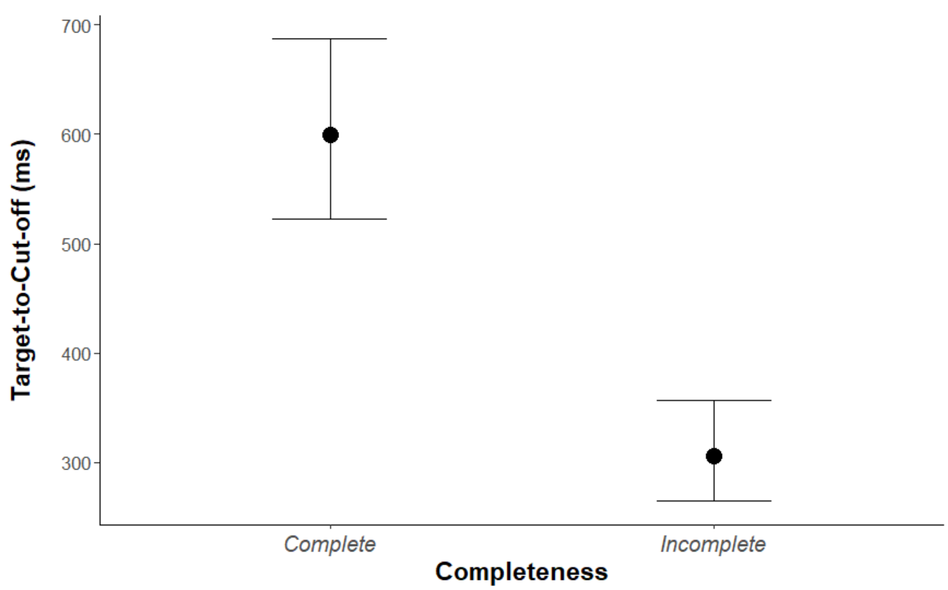
$$Percentage\ Change = \left(e^{Estimate} - 1\right) \times 100 \qquad (2)$$

Starting with *Completeness*, the analysis indicates that repairs with incomplete reparanda reach the cut-off point approximately 240 milliseconds faster than complete ones; that is, 49.3%

shorter, with a coefficient (-239.71 ms; *SE* = 28.53ms) and a 95% Credible Interval (CI) [-298.04 ms, -185.7 ms]. The 95% credible interval for the effect of *Completeness* does not include zero, suggesting a robust impact of this variable on the duration of the Target-to-cut-off, given that the effect is unlikely to be zero. Figure 12 below illustrates the effect of *Completeness* on Target-to-cut-off, as extracted from `mod1`.

Figure 12

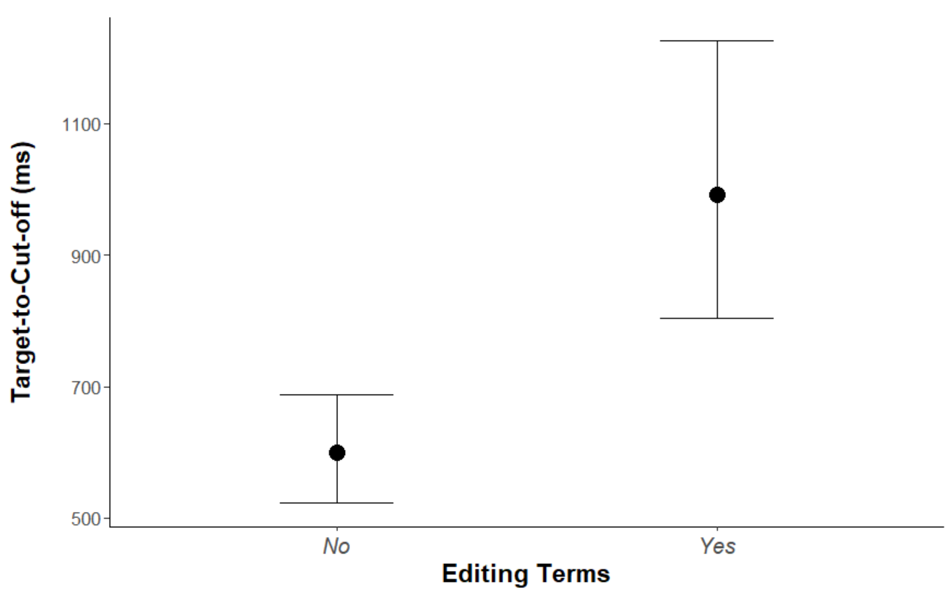
Conditional effects plot of the Log normal `mod1` *for Completeness; the error bars display 95% credible intervals; the dots represent posterior means.*



Moving on to *Editing Terms*, analyses extracted from `Mod1` reveal a relationship between the presence of editing terms and the duration of the Target-to-cut-off interval. To report on these results, it is crucial to acknowledge that, chronologically, the presence or absence of editing terms occurs after the Target-to-cut-off has ended. While this sequence of events suggests caution in interpreting these results, the findings nonetheless indicate a systematic association between *Editing Terms* and the preceding interval. It is possible that longer Target-to-cut-off durations make the use of editing terms more likely, or that both patterns are influenced by underlying properties of certain types of repairs or even different editing expressions. Despite these considerations, the observed relationship remains an interesting and robust finding. The regression coefficient shows that the presence of editing terms is associated with an increase of approximately 323 ms in Target-to-cut-off duration, representing a 65% rise (Estimate: 322.96 ms, *SE* = 86.64 ms, 95% CI [160.58, 503.53 ms]). The effect is considered strong, as the 95% credible interval lies well away from zero. Figure 13 below illustrates the association between *Editing Terms* and Target-to-cut-off as extracted from `Mod1`.

Figure 13

Conditional effects plot of the Log normal `mod1` *for Editing Terms; the error bars display 95% credible intervals; the dots represent posterior means.*



Regarding the effects of *Repair Type*, the modelling suggests that appropriateness repairs tend to be associated with higher Target-to-cut-off durations compared with error repairs. Both factual and linguistic errors showed shorter Target-to-cut-off intervals than appropriateness repairs, and the magnitude of reduction was similar across the two error types. Specifically, for both factual and linguistic errors, the interval is shorter than for appropriateness repairs by a similar measure. Factual errors had a reduction of approximately 52 ms, with a mean -51.7 ms, *SE* = 39, and a CI: [-127.51, 25.21]; likewise, linguistic errors showed a reduction of the interval, with an estimate of -66.08, *SE* = 45.08, and a CI: [-152.24, 23.9]. Both these credible intervals are wide and include a zero, which suggests that these effects could range from a substantial reduction to no change. Figure 14 illustrates the impact of the different semantic repair subtypes on Target-to-cut-off. A complete summary of estimates in milliseconds for the variables *Completeness*, *Editing Terms* and *Repair Type* in `mod1` is presented in Table 10.

Figure 14

Conditional effects plot of the Log normal `mod1` *for Repair Type; the error bars display 95% credible intervals; the dots represent posterior means.*

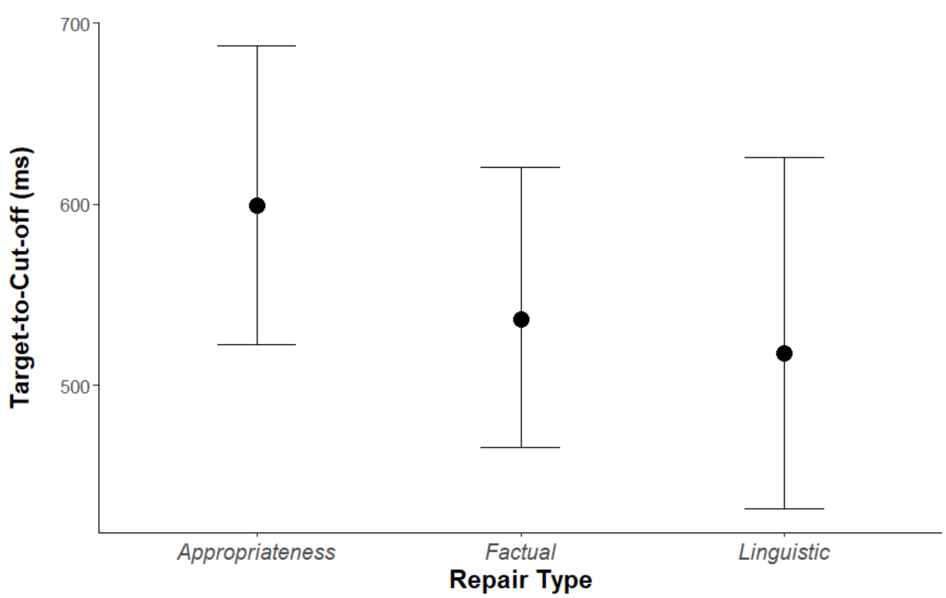


Table 10

Posterior mean, standard error, 95% credible interval and brief interpretation for each predictor in `mod1`.

	Estimate	***SE***	**Lower bound**	**Upper bound**	**Interpretation**
Intercept	490.42	33.8	428.06	560.17	*Baseline duration for complete, appropriateness repairs, without editing terms.*
Completeness Incomplete	-239.71	28.53	-298.04	-185.7	*Repairs with* incomplete *reparanda are 239.7 ms faster than complete ones.*
Editing Terms Yes	322.96	86.64	160.58	503.53	*The presence of editing terms extends the interval by approximately 323 ms.*
Repair Type Factual	-51.7	39	-127.51	25.21	*Potential reduction of 51.7 ms, but with uncertainty.*
Repair Type Linguistic	-66.08	45.08	-152.24	23.9	*Potential reduction of 66 ms, but with uncertainty.*

Note. Coefficients are back-transformed to milliseconds. Original *Log-normal* outputs are located in Appendix E.

Moving on, I will now present the effects of the interaction between the variables *Editing Terms* and *Repair Type*. Results on the relationship between *Editing Terms* and the duration of Target-to-cut-off revealed an unexpected pattern given the chronology of these two events during the repair sequence (i.e., editing terms happening after the cut-off). On the other hand, *Repair Type* alone did not show robust effects on the duration of the interval. With that in mind, `mod1` was further refined to explore the interaction between these two variables, aiming to determine whether changes in the duration of the Target-to-cut-off interval could be better explained by the interaction between *Editing Terms* and the different repair subtypes.

Results from the effects of the interaction between *Editing Terms* and the different repair subtypes revealed a clear distinction between appropriateness infelicities and linguistic and factual errors. When editing terms occur in the context of appropriateness repairs there is 65% increase in the duration of the interval, with an estimate of about 323 ms (*SE* = 86.64 ms), and a 95% CI [160.58 ms, 503.53 ms]. The effect is robust for this semantic repair subtype while for both factual and linguistic errors, the effects seem to be negligible given that results showed wide credible intervals, particularly for linguistic repairs. In more detail, factual errors had a small decrease of about 80 ms (mean -73.41 ms; *SE* = 60.49 ms, CI [182.37 ms, 64.69 ms]) while linguistic errors showed a very small increase (about half a millisecond) of the Target-to-cut-off time (mean 0.83 ms; *SE* = 119.72 ms, CI [-187.27 ms, 280.34 ms]). These findings suggest that when compared to appropriateness repairs without editing terms, the presence of editing terms can be associated to a considerable lengthening of the Target-to-cut-off interval, but for appropriateness repairs only; the magnitude of this effect is very pronounced (65% increase) and the credible interval is not close to zero, highlighting the robustness of this finding. Table 11 summarizes these results and Figure 15 illustrates the individual effects of the interaction between *Editing Terms* and *Repair Type*.

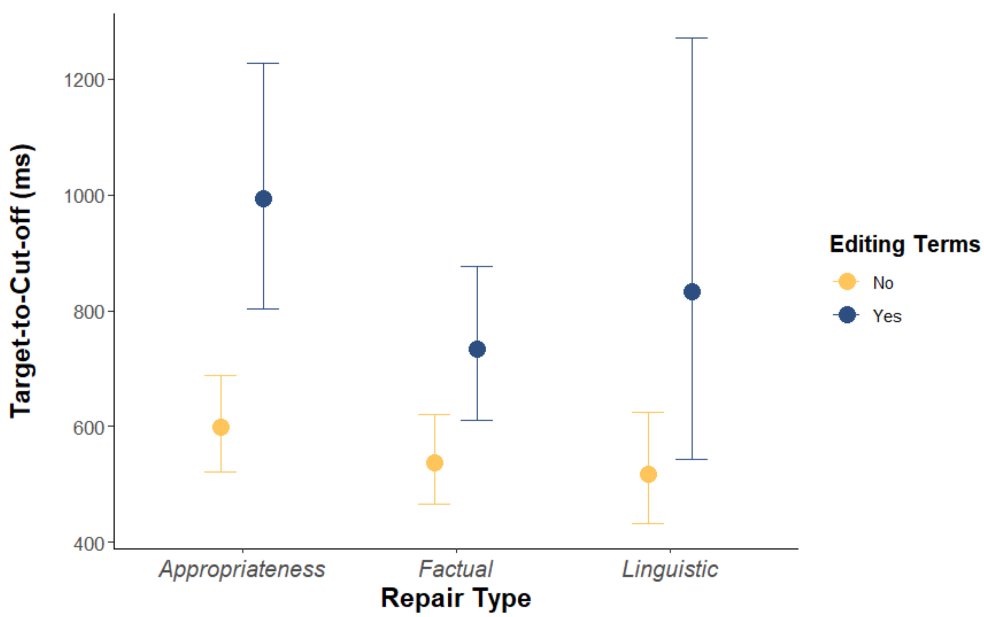
Table 11

Posterior mean, standard error, 95% credible interval and brief interpretation on the interaction between Editing Terms and Repair Type in `mod1`.

	Estimate	*SE*	Lower bound	Upper bound	Interpretation
Editing Terms Yes* Repair Type Appropriateness	322.96 ms	86.64 ms	160.58 ms	503.53 ms	*Editing Terms increase the Target-to-cut-off interval by approximately 323 ms (66% increase).*
Editing Terms Yes* Repair Type ***Factual***	-79.41 ms	60.49 ms	-182.37 ms	54.69 ms	*Editing Terms decrease the Target-to-cut-off interval by approximately 80 ms (16.2% decrease) for factual errors.*
Editing Terms Yes* Repair Type ***Linguistic***	0.83 ms	119.72 ms	-187.27 ms	280.34 ms	*Editing Terms have a negligible effect for linguistic errors, with an estimated change of 0.52 ms (0.2% increase).*

Figure 15

Conditional effects plot of the Log normal `mod1` *for the individual interactions between Editing Terms and Repair Type; the error bars display 95% credible intervals; the dots represent posterior means.*



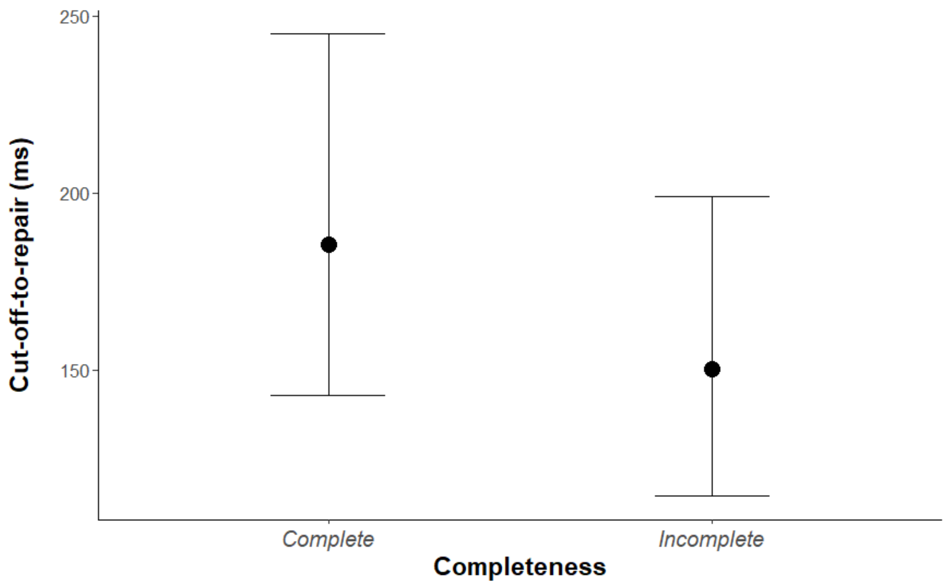
3.4.2 Relevance of Completeness, Editing Terms and Repair Type on Cut-off-to-repair

`Mod2` evaluated the effects of all independent variables on Cut-off-to-repair. Results show that only *Editing Terms* had a significant impact on the duration of this interval, while *completeness* and the variables evaluating repair semantics show small effects with credible intervals close to zero indicating no strong effects. The impact of *Editing Terms* aligns with expectations, as the time it takes to include these editing expressions within the repair stretch will inevitably increase the Cut-off-to-repair durations. After exploring significant relationships between the effects of the independent variables, neither the interactions between *Completeness* and *Repair Type* nor between *Editing Terms* and *Repair Type* showed strong interaction effects. For that reason, `mod2` was kept simple and in this section, I will only report on the results for each of the independent variables on Cut-off-to-repair, without returning to any interaction effects.

Results on *Completeness* indicate that incomplete repairs are solved 15.6 ms faster than complete ones (-15.99 ms; *SE* = 9.57 ms), representing a 14.6% reduction in the duration of the interval, though these results come with uncertainty given that the 95% credible interval crosses zero [-35,17 ms, 2.4 ms]. Figure 16 below illustrates the effect of *Completeness* on Cut-off-to-repair, as extracted from `mod2`.

Figure 16

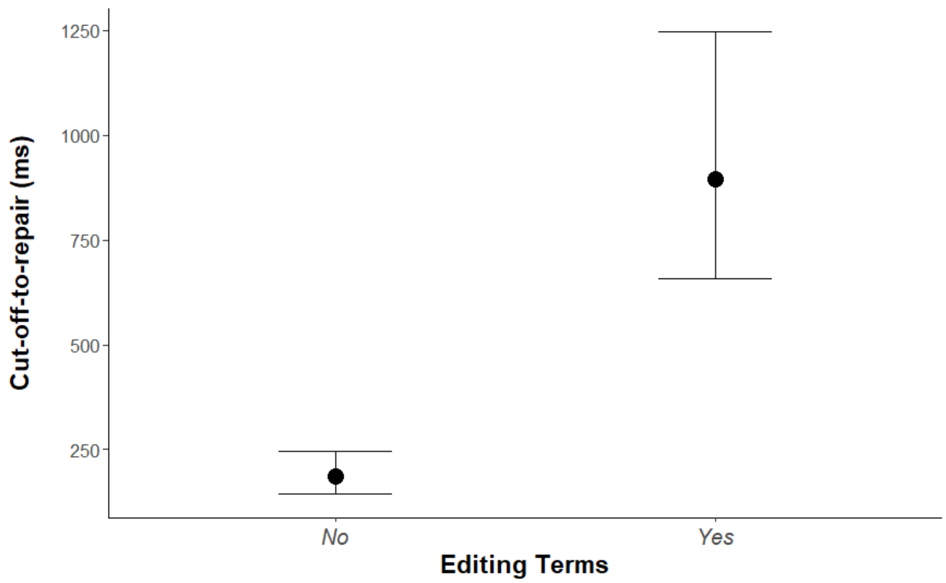
Conditional effects plot of the Log normal `mod2` *for Completeness; the error bars display 95% credible intervals; the dots represent posterior means.*



Regarding analysis of *Editing Terms*, results extracted from `mod2` reveal a strong effect of the presence of editing terms in the duration of the Cut-off-to-repair. The regression coefficient shows that the presence of editing terms increases the duration of the Cut-off-to-repair by approximately 315 ms, representing an estimated rise of 383.7% in the duration of the interval (Estimate: 315.31 ms, *SE* = 58.52 ms, 95% CI [216.33, 443.85 ms]). The impact of *Editing Terms* on Cut-off-to-repair is therefore considered robust as the 95% credible interval is nowhere near zero. These results were expected since the articulation of such editing expressions occur within the interval itself, so when speakers incorporate them in the repair sequence, it likely involves some extra time. Figure 17 below illustrates the impact of *Editing Terms* on Cut-off-to-repair, as extracted from `mod2`.

Figure 17

Conditional effects plot of the Log normal `mod2` *for Editing Terms; the error bars display 95% credible intervals; the dots represent posterior means.*



To conclude the report on the impact of the different independent variables on the Cut-off-to-repair, neither subtype of repairs (i.e., appropriateness, linguistic or factual repairs) had major impact on the duration of the interval. factual errors had an increase of around 4.11 ms, *SE* = 11.66, and a CI: [-18.73, 27.54]; oppositely, linguistic errors showed a small reduction of the interval, with an estimate of -5.77, *SE* = 14.04, and a CI: -32.82, 23.63]. Both these credible intervals include a zero, which suggests that these effects are not robust. Figure 18 illustrates the impact of the different semantic repair subtypes on Cut-off-to-repair. To conclude this section, a full summary of estimates in milliseconds for the variables *Completeness*, *Editing Terms* and *Repair Type* in `mod2` is presented in Table 12.

Figure 18

Conditional effects plot of the Log normal `mod2` *for Repair Type; the error bars display 95% credible intervals; the dots represent posterior means.*

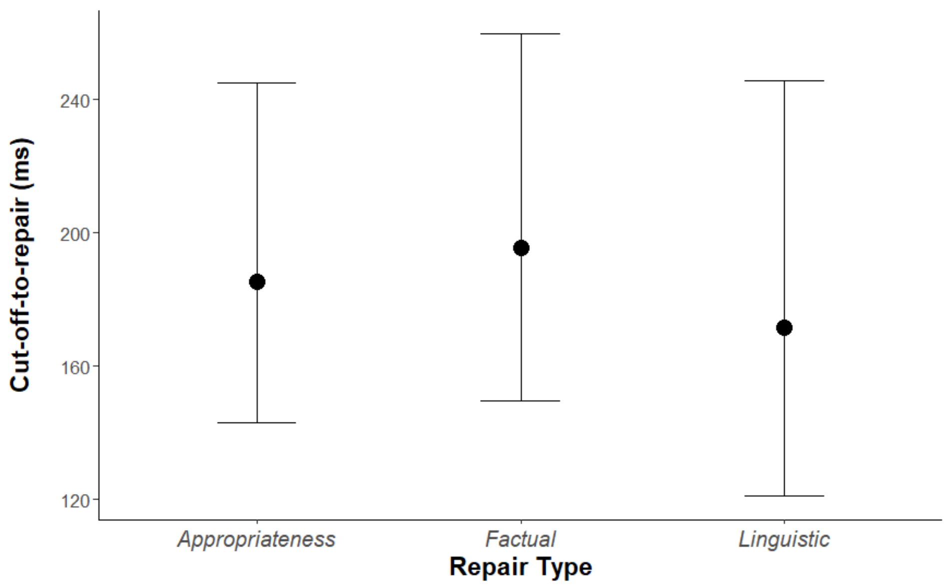


Table 12

Posterior mean, standard error, 95% credible interval and brief interpretation for each predictor in `mod2`.

	Estimate	***SE***	**Lower bound**	**Upper bound**	**Interpretation**
Intercept	82.21	10.32	64.22	104.80	*Baseline duration for complete appropriateness repairs without Editing Terms.*
Completeness Incomplete	–15.99	9.57	–35.17	2.4	*Repairs with* incomplete *reparanda are about 16 ms faster than complete ones, representing a 19.5% reduction, though the effect is uncertain.*
Editing Terms Yes	315.31	58.52	216.33	443.85	*The presence of Editing Terms adds approximately 315 ms to the interval, representing a 383.7% increase.*
Repair Type Factual	4.11	11.66	–18.73	27.54	*Factual errors increase the interval by about 4 ms (5.0% increase), though the effect is uncertain.*
Repair Type Linguistic	–5.77	14.04	–32.82	23.63	*Linguistic errors decrease the interval by about 6 ms (7.0% reduction), though the effect is uncertain.*

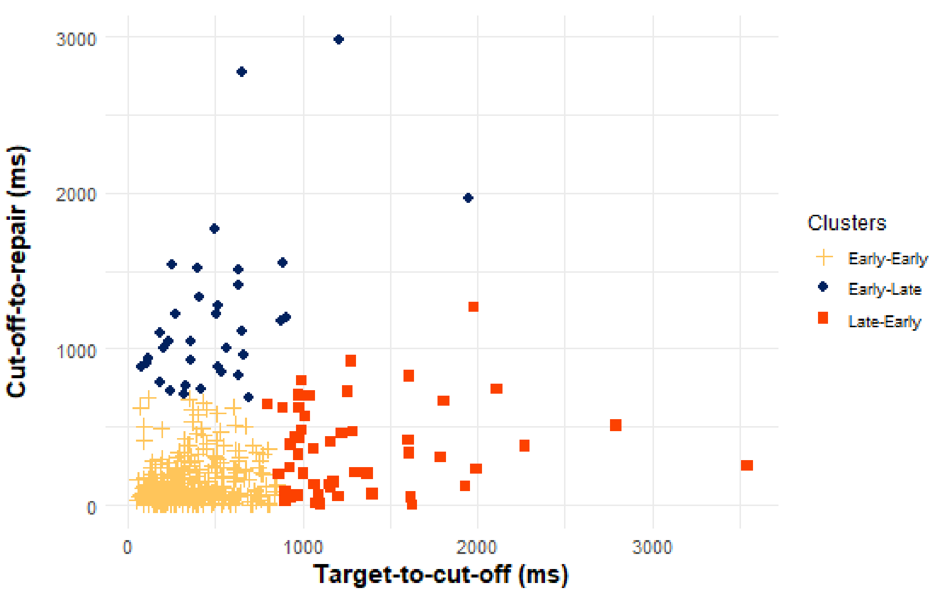
Summing up, for the Cut-off-to-repair interval only the variable *Editing Terms* was found to have a robust impact on the duration of the interval by expanding it by approximately 315 ms (37.2% rise) when editing terms are present in the repair sequence. Neither of the other independent variables nor the interactions between them showed a strong effect on the duration of the evaluated interval.

3.4.3 Relationship between Target-to-cut-off and Cut-off-to-repair

The relationship between Target-to-cut-off and Cut-off-to-repair was explored in more detail through *k-means* clustering. Results revealed three clusters as illustrated in Figure 19. First, there are repairs with a relatively short Target-to-cut-off and a relatively short Cut-off-to-repair. Following terminology introduced in earlier joint work with Plug (Vera Diettes & Plug, 2023), I call these 'Early-Early' repairs: they have an early interruption and an early repair onset. Second, there are repairs with a relatively short Target-to-cut-off but a relatively long Cut-off-to-repair. We call these 'Early-Late' repairs. Although these repairs are interrupted quite promptly, the resumption is somehow delayed, showing a late repair onset. Third, there are repairs with a relatively long Target-to-cut-off but a relatively short Cut-off-to-repair; we call these 'Late-Early' repairs. The interruption of the Target-to-cut-off comes late for these repairs but this late interruption contrasts with a short repair onset. On the possibility of a fourth cluster, Figure 19 also shows there are only a few repairs in the data set that might potentially qualify for a 'Late-Late' classification.

Figure 19

Scattergram showing Target-to-cut-off and Cut-off-to-repair durations and cluster membership.

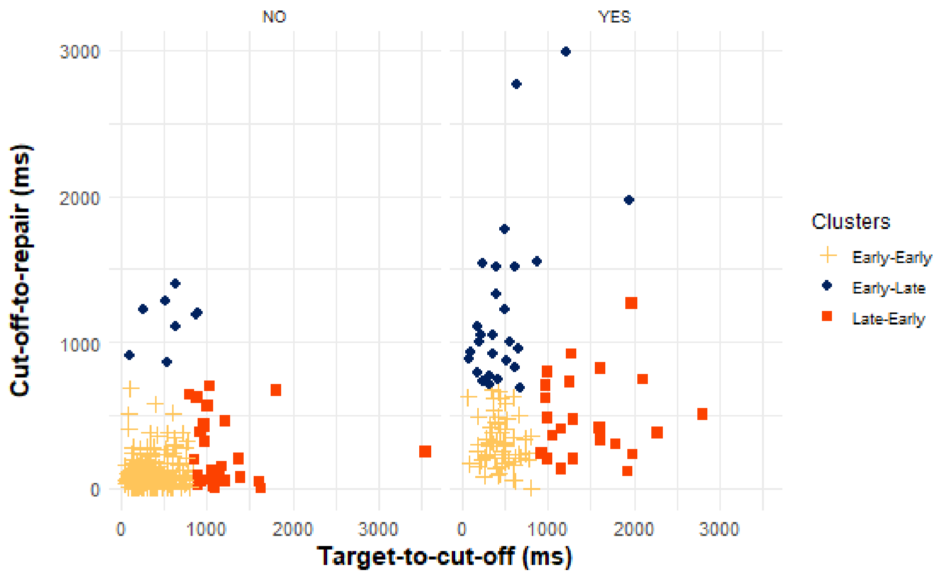


With the aim of continuing the exploration of the variable *Editing Terms*, which in our Bayesian regression showed robust effects in both Target-to-cut-off (`mod1`) and Cut-off-to-repair (`mod2`) we further split our clusters based on the presence or absence of editing expressions. As expected, there is a systematic relationship between the presence of editing terms and Cut-off-to-repair, confirming the predicted effect observed from the results of `mod2`. Interestingly, *Editing Terms* also seems systematically related to the Target-to-cut-off duration. Figure 20 shows that repairs without any editing terms predominantly fall in the Early-Early cluster, while repairs with an editing term are evenly distributed across the three clusters.

Interestingly, the presence of editing terms is not limited to repairs with delayed interruptions. Several sequences that include editing terms also exhibit short Target-to-cut-off intervals. This suggests that speakers may draw on editing terms within at least two distinct, pragmatically motivated strategies. In one, the speaker interrupts early and uses an editing term to manage the turn and create space for planning the upcoming repair. In the other, the speaker appears to delay the interruption, despite having detected the problem early, potentially continuing to speak while formulating the repair, and only cutting off when a candidate repair is available, at which point a brief editing expression may be used to introduce it.

Figure 20

Scattergram as in Figure 19, split by Editing Terms (left No*, right* Yes*)*



3.4.5 Patterns of Editing Term Use

An analysis of the distribution of editing terms across semantic repair types revealed notable differences. Factual error repairs featured the most editing term instances (*n* = 61, 61.5%), followed by appropriateness repairs (*n* = 39, 35.8%). Linguistic repairs included the fewest editing terms (*n* = 9, 8.3%). Table 13 lists the frequency of each editing term within appropriateness, factual, and linguistic repairs, with totals reflecting individual tokens across all repair sequences.

Table 13

Frequency of Editing Terms by Repair Type

Repair Type	**Count**	**Percentage**
Appropriateness	61	56%
Factual	39	35.8%
Linguistic	9	8.3%
Total	109	103%

After a closer inspection into the different editing expressions used in Colombian Spanish, the following patterns emerged. The most frequently used editing terms across repair types were *o sea* 'I mean' (*n* = 25, 22.9%), followed by the filled pause *ee* 'uh' (*n* = 16, 12.8%) (which varied in length) and lastly *bueno* 'well' (*n* = 11, 10.1%). These were followed by *pues*[29] 'well', and *o* 'or' which were used less frequently. These expressions encompass both filled pauses and discourse markers, suggesting that speakers rely on a small set of frequent terms to manage trouble in spontaneous speech. When examined by repair type, *o sea* 'I mean' was most strongly associated with factual repairs, as it appeared in approximately 50% of the cases. This editing term appears to function primarily as a clarifying or re-formulative marker. Appropriateness repairs showed a broader variety of editing expressions, including *e*, *pues* 'well', and compound markers such as *o, bueno* 'or, well' and *o, de cierta forma* 'or, in a way' reflecting a mix of timing, rhetorical, and pragmatic functions. Linguistic repairs included very few editing terms (*n* = 9, only 8.3% of the cases), and those that did occur were typically short,

[29] *Pues* is a common discourse marker in Colombian Spanish. While it can carry various meanings depending on context, in repair sequences it most often functions pragmatically as a hesitation marker or floor-holder, similar to English 'well' or 'you know'.

such as *ee* ‘uh’. This pattern supports the view that editing expressions are more likely when the speaker is managing conceptual or pragmatic complexity, whereas linguistic repairs proceed more fluently and with less need for interactional signalling. These patterns are presented in Table 14, which shows a summary of the different editing terms found, and their use, highlighting their relative frequency of occurrence across the three repair subtypes, as well as within the whole subset of repairs that included editing terms.

Table 14

Proportion of Editing Term Occurrences by Repair Type

Editing Term	**Appropriateness n (%)**	**Factual n (%)**	**Linguistic n (%)**	**Total n (%)**
o sea ‘I mean’	10 (40%)	13 (52%)	2 (8%)	25 (22.9%)
e/eee ‘uh’	5 (36%)	7 (50%)	2 (14%)	14 (12.8%)
bueno ‘well’	2 (18%)	7 (64%)	2 (18%)	11 (10.1%)
pues ‘well’ / ‘you know’	2 (40%)	2 (40%)	1 (20%)	5 (4.6%)
como ‘lke’	1 (33%)	2 (67%)	0 (0%)	3 (2,8%)
Compound markers[30]	8 (36%)	12 (55%)	2 (9%)	22 (20,2%)
Other[31]	11 (38%)	18 (62%)	0 (0%)	29 (26,6%)
Total	**39 (36%)**	**61 (56%)**	**9 (8%)**	**109**[32]

These counts reflect a tendency for editing expressions such as *eee*, *bueno* ‘well’, and *o sea* ‘I mean’ to co-occur more frequently with repairs involving conceptual or factual content, and to

[30] Compound markers refer to combinations of two or more editing expressions produced together within the same repair space, such as *o, bueno* (‘or, well’), *o sea, como que* (‘I mean, like’), *pues como (*‘well, like’*),* e, o *(*‘uh, or’*), etc.*

[31] Miscellaneous includes low-frequency and single-instance terms.

[32] Total refers to distinct editing term tokens. Some repairs included more than one marker.

be largely absent from rapid, production-based repairs. This pattern supports previous findings showing that editing terms are more likely when speakers require additional time to formulate or justify a correction (Levelt, 1983; Plug, 2016), particularly in pragmatically or factually complex utterances. Also, the patterns observed also found ground in previous descriptions showing that different subtypes of repair are associated with distinct editing term choices (e.g., James, 1972; Levelt, 1983).

These patterns are consistent with the clustering results reported in the previous section (3.4.3), which showed that editing terms were particularly prominent in the Early-Late cluster, characterized by early interruption and delayed repair onset. This suggests that editing expressions play a role in the temporal organization of repair by allowing speakers to interrupt speech flow even when the upcoming repair is not yet fully planned. Their presence is especially common in conceptually demanding repairs, such as appropriateness repairs, where speakers require additional planning time and use editing terms to manage the delay and maintain control of the turn. In this sense, editing terms contribute both to the semantic and temporal structuring of self-repair.

3.5 Discussion

In this chapter I assessed the temporal organization of lexical self-repair in Colombian Spanish unscripted speech by focusing on the distributions of Target-to-cut-off and Cut-off-to-repair intervals, and the effects a series of variables had on them. Results show that both intervals have wide duration distributions. When explored separately, Cut-off-to-repair shows indications of multimodality, whereas Target-to-cut-off does not. However, when considered together, the emergence of three clusters (i.e., Early-Early, Early-Late and Late-Early) suggests that dividing both distributions into ‘early’ and ‘late’ ranges is informative. Analyses confirmed that repairs that included editing terms have longer Cut-off-to-repair intervals than those without them; interestingly, they also have longer Target-to-cut-off intervals, which was particularly pronounced among appropriateness repairs. In consequence, the semantic subdivision between appropriateness repairs and factual and linguistic error repairs was found to influence the Target-to-cut-off in the light of the presence of editing terms. Also, as expected, repairs in which the target word is completed before the repair (i.e., repairs with complete reparanda) are associated with longer Target-to-cut-off durations than repairs in which the target word is interrupted early. These findings are discussed below with respect to the four original hypotheses guiding this study.

In relation to HYPOTHESIS A, which predicted that complete reparanda would result in longer Target-to-cut-off and Cut-off-to-repair durations than incomplete reparanda, repairs in which the target word was completed before interruption were indeed found to have longer Target-to-cut-off durations than repairs interrupted earlier. However, while completeness showed a clear relationship with Target-to-cut-off, the relationship with Cut-off-to-repair was less straightforward. In particular, Nooteboom (2010) and Nooteboom and Quené (2017) observed that Target-to-cut-off durations in elicited speech error repairs are bimodally distributed, where interrupted speech error repairs are completed quickly, while completed reparanda repairs are generally slower. In the spontaneous lexical repairs studied here, ‘early’ and ‘late’ repairs are identified, but unlike Plug (2016), I did not find that late repairs in terms of Target-to-cut-off durations, or target word completeness, are also late in terms of Cut-off-to-repair; and early repairs in terms of Target-to-cut-off are not necessarily early in terms of Cut-off-to-repair either.

Turning to HYPOTHESIS B, which addressed whether a semantic classification of repairs would predict Target-to-cut-off and Cut-off-to-repair durations, the results indicated that the semantic classification between appropriateness repairs and factual and linguistic errors did not yield substantive differences in interval durations or cluster membership. Regarding the different subtypes of repair that were studied, unlike Plug (2016), we did not observe any substantive effects of semantic repair type on interval durations or cluster membership. Plug (2016) suggests that the distinction between linguistic and factual error repairs is informative for understanding repair timing. Results do not confirm this in quantitative terms, but our observation that only a small number of linguistic errors (n = 9) contain editing terms, compared to 79 appropriateness repairs and 32 factual repairs, is consistent with linguistic and factual error repairs having distinct organizations. In total, 120 repairs included editing terms, further highlighting the rarity of their occurrence in linguistic repairs. A larger data set should allow us to explore the nature of this distinction further.

As for HYPOTHESIS C, which anticipated that the presence of editing terms would result in longer durations and would more frequently accompany appropriateness repairs, factual error repairs, and repairs with complete reparanda, the analyses robustly confirmed that editing terms have an impact on repair timing. Results on the effects of *Editing Terms* confirm that the presence or absence of editing expressions in between the cut-off and the resumption of speech is highly consequential for the temporal organization of repair, as previously informed by Levelt (1983), Nooteboom (2005b) and (Plug, 2016). They all considered editing expressions an important functional part of repair; an idea which finds ground here too. It might seem unsurprising that Cut-off-to-repair intervals which contain editing terms are on average longer than silent Cut-off-to-repairs, simply because producing an editing term takes time. However, a positive effect of the presence of editing terms was also observed on the Target-to-cut-off duration, and this effect appears to be modulated by *Repair Type*. Specifically, appropriateness repairs tended to show relatively long Target-to-cut-off intervals when editing terms were present. These instances in which editing expressions were present, were concentrated in the Late-Early cluster, characterized by delayed interruption followed by a relatively quick repair. This suggests that, particularly in conceptually demanding repairs such as appropriateness repairs (Blackmer & Mitton, 1991), speakers may strategically delay interruption, possibly to continue processing, before deploying an editing term to initiate repair rapidly. The clustering pattern highlights how editing terms may serve not only to manage the repair itself but also to reflect differences in repair planning strategies across semantic types. This pattern confirms observations by Blackmer and Mitton (1991, p. 190), who reported that conceptually-based repairs, such as those involving appropriateness issues, tend to require greater planning effort and are more frequently accompanied by editing terms. They interpreted editing terms as a strategic device that allows speakers to manage the increased cognitive demands of reformulating the utterance. Such demands may plausibly affect not only the planning phase following interruption, extending the Cut-off-to-repair interval, but also the moment of interruption itself, contributing to longer Target-to-cut-off durations in conceptually demanding repairs. Given these findings, more qualitative analysis is warranted to identify the specific subtypes of repair that have this temporal organization as well as further inspection into the use of certain types of editing expressions.

Finally, HYPOTHESIS D predicted that the distributions of the intervals would show durations close to 0 ms as well as around 1,000 ms, suggesting indications of bimodality or multimodality. Results revealed a wide range of interval durations, with Cut-off-to-repair intervals found as short as 1 ms and Target-to-cut-off intervals starting from 50 ms, and with instances in both intervals exceeding 3,000 ms. This finding aligns with the predictions by Nooteboom (2010) and Nooteboom and Quené (2017, 2019, p. 54) regarding the expected dispersion of repair timing associated with early and late error detection. Although the mean durations were notably shorter than 1,000 ms, the broad distributions, particularly the presence of intervals well beyond that threshold, support Nooteboom and Quené's (2019, p. 54) and Quené and Nooteboom's (2024, p. 4) suggestion that spontaneous lexical repairs may exhibit even longer durations than those typically found in elicited phonological repairs, reflecting the increased cognitive complexity and variability of spontaneous speech monitoring.

The found clustering structure observed in the data provides additional support for HYPOTHESIS D, which also predicted that multimodal distributions might emerge more clearly when considering Target-to-cut-off and Cut-off-to-repair jointly. When examined separately, Cut-off-to-repair distribution showed only indications of multimodality while Target-to-cut-off did not. However, when both intervals were analysed together, three meaningful clusters emerged: Early-Early, Early-Late, and Late-Early. The Early-Early cluster reflects cases in which speakers interrupt quickly and repair immediately, a pattern consistent with previous accounts of early error detection through inner speech monitoring (Hartsuiker & Kolk, 2001; Levelt, 1983, 1989) and with empirical findings on low-complexity or production-based repairs (Blackmer & Mitton, 1991; Nooteboom, 2010; Plug & Carter, 2014). This type of timing may reflect highly automated monitoring and repair processes. In contrast, the absence of a Late-Late cluster and the presence of a Late-Early cluster fit with the idea that speakers can delay interruption to maximize fluency (Blackmer & Mitton, 1991; Nooteboom & Quené, 2019; Seyfeddinipur et al., 2008). More specifically, while both intervals were expected to show wide distributions compatible with early and late timing combinations (e.g., Nooteboom, 2010), the lack of a Late-Late configuration suggests that detection during overt speech does not always result in prolonged Cut-off-to-repair intervals. This absence underscores the importance of considering strategic delay and not merely late detection *per se*, but the intentional postponement of interruption or of repair initiation (Nooteboom & Quené, 2019; Quené & Nooteboom, 2024; Seyfeddinipur et al., 2008). The presence of both Late-Early and Early-Late clusters reflects two distinct delay strategies: one where speakers delay the interruption itself (Late-Early), and another where they interrupt early but delay the repair through the use of editing expressions (Early-Late). Seyfeddinipur et al. (2008) suggests that the use of editing expressions might be an alternative 'repair delay' strategy to delaying the cut-off: after an early interruption, speakers can maintain fluency by using editing terms while planning the repair. In fact, Figure 20 shows that this accounts for most instances in the Early-Late cluster, as most contain one or more editing terms. This finding also appears to support the idea that speakers use editing terms to let the listener know they are having issues with their speech (Seyfeddinipur et al., 2008, p. 841). Taken together, these findings suggest that *delay* is not a passive outcome of late detection, but a pragmatically motivated choice, deployed flexibly depending on interactional goals. Further work is needed to expand this data set, so that the robustness of the Late-Early and Early-Late temporal organizations can be evaluated, as well as the absence of a commonly observable Late-Late one.

In summary, the results partially supported the original hypotheses. HYPOTHESIS A found confirmation in the relationship between completeness and Target-to-cut-off duration, although Cut-off-to-repair did not systematically follow. Hypothesis B received less support, as semantic repair type did not robustly predict timing differences. Hypothesis C was confirmed: the presence of editing terms significantly influenced both intervals. Finally, Hypothesis D found partial support, with Cut-off-to-repair showing indications of multimodality and clustering analysis revealing meaningful patterns. These findings highlight the complex and dynamic nature of repair timing in spontaneous speech.

Chapter 4

Prosodic Marking in Lexical Self-repair

4.1 Introduction

This chapter is concerned with prosodic marking in the context of repair. As explained by Cutler (1983), once faced with troubles in their speech, speakers have a range of options. The first one is whether to repair their error or infelicity. If a repair emerges, then comes the second choice, which is whether, consciously or not, speakers enhance their repairs by adding salience to their repair productions. Oppositely, speakers can minimize the disruptive effect of the trouble source by not adding any sort of salience to corrected speech material, therefore, not MARKING their repairs (Cutler, 1983, p. 80).

One of the matters that has surrounded the study of repair's prosody is how systematic speakers' behaviour towards marking is. The question remains far from solved and while trying to answer this question, researchers have pointed out to a range of constraints affecting speakers' behaviour towards marking in repair; therefore, the aim of the investigation reported in this chapter is to shed light on the role of prosodic marking in repair processes. More specifically, this research is devoted to investigating prosodic marking in lexical self-repair in unscripted Colombian Spanish to help understand whether the same constrains towards marking are shared across languages. As in Chapter 3, I evaluate how the SEMANTICS of the repair sequence, reparanda COMPLETENESS and the presence or absence of EDITING TERMS affect and interact with the prosody of the repaired speech. In this occasion, I tested whether these variables, together or individually, influenced the prosody of the repair sequence. I did so by comparing acoustic information on PERIODIC ENERGY, PITCH and SPEECH RATE of reparanda to those of the repairs.

These acoustic parameters, which I will explain in more detail in the following sections of the chapter, were selected as phonetic correlates of prosody following work on the study of prosodic marking in repair (e.g., Nooteboom, 2010; Plug, 2014; Plug & Carter, 2013), while also considering new approaches to prosodic prominence based on periodic energy and pitch intelligibility (Albert, 2023; Albert et al., 2018; Albert & Nicenboim, 2022). I used Bayesian regression for modelling results and, similarly to the previous report on temporal organization, the data for the analyses presented here was extracted from the UCSI corpus.

As I have illustrated in earlier sections (1.2.1 and 3.1), in lexical self-repair, a speaker rejects one lexical choice in favour of another, as in (18) below, *El* ***Esta-****; eee, La* ***República de Colombia***, in English, 'The **Sta-;** eeeh, The **Republic of Colombia**'. In this example, the speaker rejects the more generic noun *Estado* 'state', for the name 'Republic of Colombia'.

(20) Appropriateness self-initiated, self-repair

Spanish (Vera Diettes sp_17_18_conv_kv_part28)

01	A:	El **Esta-;** eee, La **República de Colombia.**
		'The ***Sta-;*** *eeeh,* The ***Republic of Colombia***'.

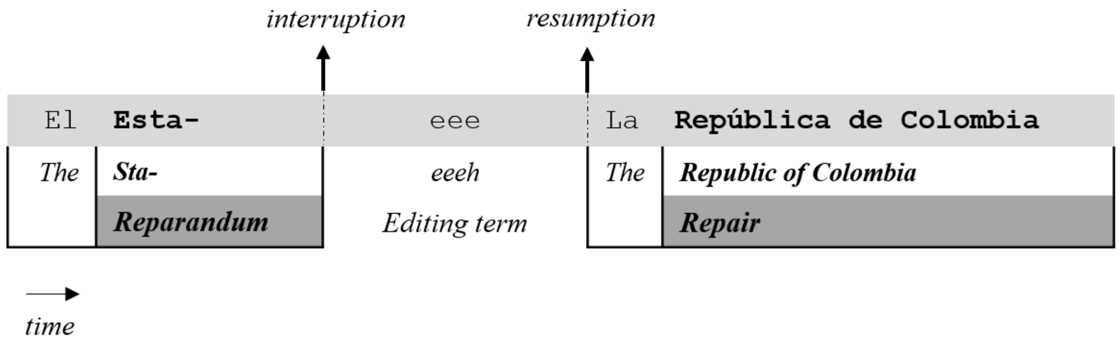
I focus on the acoustic features of both the reparandum and the repair, in (20) , the incomplete realization of the noun 'State' and its replacement in the repair. Editing terms and other surrounding speech (e.g., preceding or repeated words) are excluded from the acoustic analysis. This selection procedure is explained in more detail in the methods section on phonetic data extraction (4.3.5). I studied prosodic marking by observing three acoustic parameters that can be associated with prosodic prominence, PERIODIC ENERGY, PITCH and SPEECH RATE. Periodic energy is defined as ‘a measurement of the acoustic power of periodic components in the signal’ (Albert, 2023, p. 55) that can be associated to the prosodic strength of the speech. In this research has been chosen as an alternative to the commonly used intensity measurements in decibels that are associated with loudness. Pitch refers to that aspect of sound that allows us to classify speech signals as *high* or *low* by using our ears, and which is associated with the actual rates of vibration or frequency of the sound (i.e., fundamental frequency) (Ladefoged, 2017, p. 22). Lastly, speech rate refers to the speed at which segments and syllables are articulated, and which is measured as a count of units per second (i.e., segments or syllables per second).

To study prosodic prominence in repair (i.e., repair prosodic marking), I followed the model proposed by Albert (2023, p. 142), which holds that ‘the interaction between fundamental frequency (F0) and periodic energy should lead to more comprehensive representations of pitch in speech and beyond’. For extracting the data for the analysis of prosodic prominence in repair, I used, on the one hand, the open-source ProPer Workflow (Albert et al., 2024), a toolbox that extracts measurements based on the acoustic periodic energy to study prominence in a quantifiable manner. Employing ProPer allowed for the extraction of measurements of F0 to study pitch, and measurements of periodic energy to analyse the strength of the relevant pieces of speech, based on the power of periodic components in the signal. In addition, measurements of speech rate were computed by obtaining counts of articulation rate of segments and syllables per second. By observing periodic energy, pitch and speech rate on both the reparandum and the repair, I obtained a comprehensive view of the prosodic shape of the relevant items in the repair sequence, which makes it possible to determine whether there is preference for prosodic marking in repair in Colombian Spanish and also, if any of the proposed predictors can be linked to marking. By completing the analyses proposed for this chapter, I present a full picture of the phonetics of all relevant items within the repair sequences, beginning at the start of the reparandum and reaching up to the end of the repaired speech, complementing the temporal analysis of the Target-to-cut-off and Cut-off-to-repair intervals presented earlier in Chapter 3.

Earlier in section 1.3.3, key research on prosodic marking was presented as an introduction to this thesis. I will now go in more depth presenting all relevant work to the study of prosodic marking in self-repair. Most of the knowledge we currently possess on the matter derives from the investigations by Cutler (1983) and Levelt and Cutler (1983), who were the first to establish the concept of prosodic marking on repair. The analyses I present on this chapter also build on their work but more specifically, on recent research by Plug (2011, 2014, 2015a, 2016) who investigated the phenomenon through qualitative and quantitative assessments, including acoustic, auditory and discursive analysis of prosodic marking in repair. I will present these and other papers in detail at the literature review section of the chapter (4.2). For the analyses, I focus on the speech at the REPARANDUM, as repairable piece of speech, and the REPAIR as preferred lexical choice. Note that presence or absence of editing terms is factored as variable within the analysis, but the actual speech deployed for the production of editing terms is not analysed phonetically. Figure 21 illustrates this schematically in relation to time. There, we can observe the two pieces of speech from which the phonetic information is extracted; that is, in the example, the REPARANDUM, *Esta-* 'Sta-' and the REPAIR, *República de Colombia* 'Republic of Colombia'. Both the editing term 'eeeh' in between reparandum and repair, and the article *El* 'The', preceding the reparandum (repeated at resumption), are omitted. More details on the segmentation and preparation of the relevant pieces of speech for analysis will be given in the methods section of the chapter (4.3).

Figure 21

Schematic representation of the Reparandum and the Repair in (20).



In this chapter, prosodic marking is investigated in the light of repair semantics, reparandum completeness and the presence or absence of editing terms. Let us recall that the semantic subdivision splits repairs into issues of appropriateness versus linguistic and factual errors; the former refers to constraints related to pragmatic felicity, as in *The State* in contrast to *The Republic*, seen in (20) above, while the latter refers to either linguistic or factual errors. A linguistic error is, for example, using the noun *aggravation* instead of the verb *worsen* in the sentence 'how does this *aggravation* someone's life?'. In contrast, factual errors are those like '*savage* animals' instead of '*wild* animals'. *Completeness* refers to the distinction between complete (i.e., fully articulated) and incomplete (i.e., interrupted) reparanda. For example, in (20) the interrupted articulation of the word *Estado* ('State'), partially articulated as *Esta-*, is an example of an incomplete reparandum. (See (1) and (2) for further illustration on this contrast). *Editing Terms* is concerned with the presence interjections or hesitation markers such as 'uh', 'I mean' or 'eeeh', in between the reparandum and repair, as shown in (20). By exploring these variables, this chapter reports on the prosodic characteristics of both the reparandum and the repair with the aim of observing how different they are phonetically and to determine whether any contrast between them hints towards marking on the repair.

The present study addresses the question on whether the semantic subtype of a repair, such as linguistic, factual, or appropriateness repairs, influences its prosodic realizations in spontaneous speech. It also examines whether reparandum completeness, used here as an index of detection timing, weighs in on the prosodic marking of repair, potentially offering insight into processes of speech monitoring during spontaneous self-repair. Specifically, the comparison between incomplete reparanda (indicative of early detection in inner speech) and complete reparanda (associated with late detection in overt speech) allows for an evaluation of how detection timing may shape the phonetic prosodic characteristics of repair. Furthermore, by analysing the use of editing terms alongside prosodic cues, the study explores whether these two signalling strategies function independently or in complementary ways when speakers repair.

More broadly, this chapter and thesis represent an effort to extend the work initiated by Plug (2011, 2014, 2016) and Plug and Carter (2013) on the phonetics of spontaneous repair, which remains a significant gap in the literature. While Nooteboom's influential account of early versus late detection and the prosodic marking of repair has largely been developed through the study of experimentally elicited errors, Plug's and Plug and colleagues' work stands out as one of the few attempts to test these claims in spontaneous data. Building on this line of research, the present analysis offers new evidence from conversational Colombian Spanish, advancing our understanding of how detection status, semantic repair type, and editing expressions interact with prosodic marking in naturally occurring speech and cross-linguistically.

The rest the chapter has the following structure: Section 4.2 presents all relevant discussions around the concept of prosodic marking in repair, as well as a review of previous findings in languages different from Spanish. Section 4.3 (Method) is divided into five sub-sections. There, I first describe the specifics of the data set used for prosodic analysis (Section 4.3.1). Then, I explain the segmentation process and how I extracted the individual reparandum and repair items out of every repair sequence for phonetic analysis (Section 4.3.2). Subsequently, I present the procedure for obtaining periodic energy data from ProPer together with its derived measures of F0 in Hertz (Hz) for the analysis pitch, as well as the process for extracting speech rate measurements (Section 4.3.5). In the following section (4.3.3), I present the items classification, which allowed me to transform them into individualized predictors. The last part of the method section (4.3.5) outlines the details of the statistical models used for conducting the Bayesian regressions. Later, Section 4.4 presents the results and shows the outputs from the statistical models. Lastly, Section 4.5 discusses the implication of the reported findings for our understanding of prosodic marking in repair based on the findings for Colombian Spanish.

4.2 The Prosody of Repair

This section focuses on providing a review of relevant work that can be related to the analysis of prosodic marking in repair. This account builds up from a broad scope and includes studies that have explored prosodic marking on various types of repairs and in languages different from Spanish. Similarly to the report on temporal organization on Chapter 3, in this review I included papers that go beyond the focus of this investigation, which is lexical self-repair. The reason for this is because it is equally worth finding whether previous results on prosodic marking in other languages and other repair subtypes are generalizable to lexical repairs in spontaneous Spanish.

4.2.1 Multi-parametric Perspectives on Prosodic Marking

Prosodic marking refers to a speaker's behaviour that makes certain repairs noticeably different, or marked, in comparison to the original utterance, or reparandum (Cutler, 1983). As Cutler (1983, p. 84) explains, such marking 'can be realised in several different ways, by longer relative duration, noticeably higher or lower pitch, higher or lower amplitude, or a combination of pitch, amplitude and durational effects'. One can think that after an error or infelicity has emerged, speakers would make the repair more prominent than the error so that the correct choice is highlighted over the troubled realization. However, there is evidence pointing to diverse results in the way speakers deal with their repairs and the reasons that motivate speakers to mark their corrections (e.g., Goffman, 1981; Levelt & Cutler, 1983; Nooteboom, 1980, 2010; Plug, 2014; Plug, 2015a, 2016; Plug & Carter, 2013).

Let us start by defining what the differences between marked and unmarked repairs are, and the implications of such distinction. Levelt and Cutler (1983, p. 206) explain that for UNMARKED repairs, speakers produce a correction that mimics, as much as possible, the same pitch, amplitude and relative duration of the reparandum, while MARKED repairs contrast to the reparandum. This implies that there is not necessarily a direct relationship between higher pitch, amplitude or extended duration and marking, and vice versa (e.g., lowering pitch or reducing energy is usually associated with an unmarked repair). Thus, marking can be achieved by either increasing or decreasing pitch, amplitude or relative duration (Levelt & Cutler, 1983, p. 206). As explained by Cutler (1983) patterns of error correction can be linked to the nature of the repair itself and to the context in which the repair is embedded. Research on repair prosody have already looked at some of these factors. We have, for example, explorations on semantics (Levelt & Cutler, 1983; Plug, 2011, 2015b, 2016; Plug & Carter, 2013), the moment of interruption and its relation to speech monitoring (Carter & Plug, 2011; Levelt & Cutler, 1983; Nooteboom, 2010; Plug & Carter, 2013), informational redundancy (Plug, 2011), frequency related factors, frequency related factors (Plug, 2015b, 2016; Plug & Carter, 2013), discourse constraints (Plug, 2015a) and, in more general terms, what is considered as 'marking' in repair and how it is conceived functionally by speakers and, in more general terms, what is considered as 'marking' in repair and how it is conceived functionally by speakers (Cutler, 1983; Goffman, 1981; Plug, 2014).

Plug (2014) explains that prosodic marking in self-repair is distinguished from phonetic 'upgrade'[33] by explaining that while marking is often associated with higher pitch and intensity, it is not frequently linked to increased speech tempo (Plug, 2014, pp. 12-13). On the contrary, faster articulation may 'downgrade' a repair rather than making it more salient. Following Plug's (2014) approach to marking, this investigation adopts a multi-parametric phonetic analysis, drawing on pitch, intensity, and speech rate. However, unlike previous work, this study also includes periodic energy as an acoustic correlate, used here as an alternative to intensity, to examine prosodic marking. Most previous studies have focused on pitch and intensity jointly (e.g., Nooteboom, 2010; Plug & Carter, 2013) as they are commonly correlated, while some others have incorporated relative speech rate measures to assess prosodic features in repair (Plug, 2011, 2016). Yet, to my knowledge, no existing study has reported on the use of periodic energy in this context. As this analysis examines prosodic marking via periodic energy, pitch, and intensity, the following sections will review prior work investigating pitch and intensity and

---

[33] See Plug (2014) for details on the phonetic 'upgradring-downgranding continuum' in relation to prosodic marking.

speech rate, in order to assess how marking in repair behaves across these parameters and in the light of factors such as semantics, editing terms and completeness. Additionally, this chapter will consider whether periodic energy offers a viable alternative to intensity in accounting for prosodic strength in the context of repair.

4.2.2 The Role of Repair Semantics in Prosodic Marking

Having established the basis for a multi-parametric approach to prosodic marking, this section will focus on how the semantic nature of the repairs might influence the choice to mark it prosodically. Previous work has investigated whether speakers treat factual corrections differently from appropriateness-based adjustments, and how these distinctions has been observed in prosodic behaviour.

A key starting point for understanding how semantics influence prosodic marking in self-repair lies in the contrast between early theoretical accounts and more recent phonetic investigations. Cutler's (1983) and Levelt and Cutler's (1983) foundational work framed prosodic marking as a listener-oriented strategy. They proposed that errors, which require the correction of incorrect information, are more likely to be marked than appropriateness repairs, which serve to elaborate rather than to contrast information. In this view, marking appears to signal a change from incorrect to correct information, helping the listener distinguish between what was misspoken and what was actually meant. However, more recent corpus-based research by Plug (2011, 2014, 2015a, 2016) complicates this picture. Drawing from spontaneous speech data and detailed acoustic analysis, Plug argues that prosodic marking is far less systematic than previously thought and appears to be mediated by pragmatic and discourse-related factors, including speaker orientation, conversational context, and cognitive effort. This perspective invites a re-examination, ideally cross-linguistically, of how repair semantics shape prosodic choices during self-repair . In the following paragraphs I will go through both Levelt and Cutler's and Plug's work with the aim of reviewing both perspectives in more detail.

Cutler's (1983) started work on the function of prosody by informing on 'a large number of tape-recorded errors'. Auditory analysis indicated that unmarked corrections were either phonetic or lexical, but marked corrections only happened in lexical errors, with marking not being norm in their data; only 32% of lexical errors were found to be prosodically marked (Cutler, 1983, pp. 83-84). Following these results, and in an attempt to find out the reason for the large number of unmarked repairs, Levelt and Cutler (1983) explored the moment of interruption with focusing on the semantic dimension of repair by observing the distinction between appropriateness and error repairs. Recall that error repairs are those in which the troubled item is wrong and 'must be replaced by a correct version of the intended message' whereas in appropriateness repairs, the trouble item is not the most suiting choice for the context (Levelt & Cutler, 1983, p. 207). The authors performed independent judgements on a collection of 412 lexical repairs looking for intonation marking. Their aim was to determine whether the prosody of the repair was the same, or different from that of the reparandum by looking for variations in pitch, amplitude and duration. After comparing their judgements, only those repair items for which there was agreement on the prosodic classification were retained, resulting in a final sample of 299 cases (Levelt & Cutler, 1983, pp. 208-209). Within this set, 53% of error repairs were classified as marked, compared to just 19% of appropriateness repairs (p. 212). The authors explained that the act of correcting an error was motivated by the need to *contrast* 'correct' to 'incorrect' information, while for appropriateness issues, repairing was an act of *elaboration* on what has been said rather than a contrast between two different information. Findings then provide evidence to support the idea that speakers' preference for marking responded to the need to contrast information, adding that error repairs more often required marking, while appropriateness issues needed elaboration rather than contrast (Levelt & Cutler, 1983, p. 212). This claim aligns with Lindblom's (1996) proposal holding that new, important information is more likely to be produced with prosodic emphasis than old, unimportant material. Because error repairs introduce new information, corrections in this context are treated as communicatively important and directed at the listener. According to Levelt and Cutler (1983) this type of information is more likely to be prosodically marked, as it functions to contrast new content with what was previously said.

Subsequent studies, however, have complicated this view. As introduced earlier, Plug and collaborators advanced research based on spontaneous Dutch data, applying acoustic methods to test the relationship between repair semantics and prosodic marking. Plug's (2011) work on information redundancy looked at phonetic reduction in a collection of self-initiated, self-repairs of different nature. Plug also observed speech tempo as an acoustic parameter within the picture of prosody in repair studies. Tempo encompasses the well-known concepts of 'reduced' and 'expanded' speech. According to the H&H theory (Lindblom, 1990), speakers adapt their production to the perceptual needs of the listeners and in relation to the contextual information available, thus varying their production on a continuum between *hyper*- and *hypo*-speech. Reduced, hypo-articulated or HYPOSPEECH, are those instances in which segments have been articulated fast, and are, therefore reduced. Oppositely, 'expanded', hyper-articulated or HYPERSPEECH corresponds to those instances produced with no time pressure, resulting in longer individual segments and, therefore, slow speech (Lindblom, 1990, p. 404). Findings indicated that articulation rate of repairs increased, leading to hypo-articulation, regardless of whether the repair addressed an error or an appropriateness issue. The latter challenged the idea that semantic type alone could determine the prosodic behaviour by speakers at least when based on tempo. Recall that Levelt and Cutler's (1983) semantic analysis derived from task-oriented dialogues, while Plug's (2011) data came from free conversations. Thus, Plug (2011, p. 296) wonders whether the difference between errors and infelicities 'is more significant in the context of an instruction that in the context of an informing free conversation'. Secondly, he also draws attention to the possibility that speakers solve the issue as fast as possible, therefore, hypo-articulate, as a pragmatically motivated face-saving strategy after an error has occurred, which is consistent with the identified pattern of reduction in repair and similar to Goffman's (1981) speaker-oriented analysis of repairs that cause embarrassment, view that I will discuss in more detail in Section 4.2.3.

Building on this, Plug and Carter (2013) investigated prosodic marking in 216 spontaneous Dutch self-repairs, analysing pitch and intensity through both acoustic and auditory methods. They attempted to explain the observed frequent hypo-articulation in repairs earlier observed by Plug (2011) and questioned whether the low incidence of prosodic marking might reflect the nature of spontaneous speech rather than a lack of emphasis. They suggested that the lower marking rates observed in spontaneous speech, compared to Levelt and Cutler's (1983) task-based data, may reflect the more pragmatic nature of everyday conversations. Their sample included 129 error repairs and 87 appropriateness repairs, with errors further divided into factual and linguistic subtypes to explore semantic effects. Results showed low overall marking rates, and only a weak tendency for errors to be marked more than appropriateness repairs, a pattern that only emerged when factual errors were distinguished from linguistic ones. The acoustic data supported the auditory judgements. Based on the findings, the authors emphasized the importance of distinguishing between elicited and spontaneous contexts when interpreting prosodic marking.

Plug (2014) extended this line of work by incorporating tempo, this time as a third prosodic dimension alongside pitch and intensity. He analysed 548 lexical and phonological repairs extracted from Oostdijk's (2002) spoken Dutch corpus, and confirmed that pitch and intensity increases were the strongest predictors of marking. However, when it came to speech rate, results indicated that even marked repairs were typically produced faster than their reparanda, a pattern of hypo-articulation consistent with prior findings (Plug, 2011). Crucially, although both marked and unmarked repairs were produced at a higher tempo than their reparanda, unmarked repairs exhibited a greater degree of acceleration. This suggests that while prosodic marking involves pitch and intensity increases (i.e., phonetic upgrading), it often co-occurs with reduced articulation effort. This pattern supports the idea that speakers may emphasize corrections while moving past errors swiftly, a face-saving strategy that aligns with Plug's (2011) interpretation of prosodic marking as pragmatically motivated. Notably, hypo-articulation does not constitute an 'upgrade' in phonetic terms, as it entails neither increased duration nor articulatory effort. Methodologically, Plug (2014) adopted a similar approach to that of Plug and Carter (2013) for pitch and intensity, but with a key difference: auditory judgements that had been classified as 'possibly marked' were grouped with 'unmarked', thus isolating only clearly marked cases. All things considered, Plug (2014) proposes that prosodic marking in self-repair 'is a compromise between emphasising the corrected talk' (i.e., by increasing pitch and intensity of the repair) and 'getting the job of correcting as quickly as possible' (i.e., by hypo-articulating the repair) (Plug, 2014, p. 13).

In addition to refining the semantic categories, Plug (2015a, 2015b) also began to question their explanatory power. While these studies still acknowledge distinctions such as error versus. appropriateness, they highlight that prosodic marking may not align neatly with these categories. Instead, Plug (2015a) proposes, after analysing instances of lexical replacement self-repair, (i.e., cases in where one word is replaced by another) that factors such as information value, speaker perspective, and discourse salience play a critical role in shaping whether and how marking occurs. These observations suggest that repair semantics alone cannot account for prosodic behaviour, and that broader contextual and interactional factors may be required to explain the patterns observed, a point explored further in Section 4.2.3, which focuses on reviewing work looking at listeners' orientation and discursive factors in repair.

Lastly, Plug (2016)[34] completed an extensive quantitative analysis of instances of lexical self-repair. There, the effects of semantics on speech tempo were again examined by investigating the distinction between appropriateness and factual errors, but, extending the semantic framework by further dividing errors into factual and linguistic repairs, following suggestions by Plug and Carter (2013) and also by Plug (2015a), who highlighted the need for more context-sensitive analysis of self-repair. In a comparison of the speed of both the repair and reparandum, findings indicated that the articulation rate of the repair stretch is above that of the corresponding reparandum, which accounts for the previously reported results on reduction (i.e., hypo-articulation) on spontaneous self-repair (Plug, 2011).

---

[34] Apart from reporting on temporal organization, as discussed in Sections 3.2.1 and 3.2.2.

In sum, the contrast between Levelt and Cutler's (1983) listener-oriented account and Plug's (2011, 2014, 2015a, 2016) corpus-based findings reflects a shift in how prosodic marking is understood. While Levelt and Cutler view marking as a tool for clarifying errors, Plug's work presents a more context-sensitive perspective, where marking appears to be less systematic and shaped by pragmatic considerations. When it does occur, it typically involves increased pitch and intensity alongside reduced articulation rate, which functions as a strategy that balances emphasis with efficiency. Rather than a binary distinction, prosodic marking emerges as a compromise between signalling correction and moving past the error fast (Plug, 2014, p. 13), aligning with broader models that highlight the interplay of informational, interactional, and phonetic factors.

4.2.3 Listeners' Orientation and Discourse Context

Approaches to prosodic marking in repair have long differed in their orientation and methodological scope. One of the earliest speaker-oriented approaches was that of Goffman (1981), who focused on the discourse context in which the repair occurred. In his qualitative study of naturalistic speech from English radio broadcasts, Goffman (1981) classified repairs as either 'flat' (i.e., unmarked) or 'strident' (i.e., marked) based on their salience, with the aim of capturing how prominently speakers marked their repairs in response to social or interactional pressures. Results revealed that the distinction between repairs being 'flat' or 'strident' arose from speakers 'level of unease, or embarrassment'. He suggested that *flat* repairs occur in situations where there is no reason for embarrassment; these repairs are more subtle and less disruptive, smoothly fixing the error without drawing attention to it. Conversely, *strident* repairs are more pronounced and noticeable, which diverts attention away from the trouble source

(Goffman, 1981, pp. 215-216). Under this view, speakers' decisions on marking are based on how their image is being projected at the time of speaking and producing errors.

Building on this perspective, Plug (2015a) examined 274 spontaneous lexical repairs, 74 marked and 173 as unmarked, which were classified using previously established auditory judgements (See Plug, 2014) to classify 74 as marked and 173 as unmarked. While earlier studies had focused primarily on semantic repair types, Plug (2015a) proposed that prosodic marking is better understood as context-sensitive and influenced by the speaker's communicative goals and the discourse environment. Drawing on Sanders' (2005, p. 57) account of discourse as shaped by speaker assumptions about knowledge, motivation, and competence, Plug examined how face-threatening situations and information salience correlate with marking. Plug's analysis identified three discourse contexts consistently associated with prosodic marking: maintaining discourse coherence, maintaining epistemic authority, and strengthening or weakening speaker commitment (Plug, 2015a, p. 87). Marking was more frequent in repairs that addressed ambiguities, situations where clarity was crucial for the listener. This aligns with Levelt and Cutler's (1983) listener-oriented account, where prosody serves to highlight relevant information. In addition, Plug observed that marking also occurred in socially sensitive contexts, where speakers might risk losing epistemic authority or credibility. In such cases, prosodic marking may function as a face-saving strategy, emphasizing the correction and reasserting the speaker's stance, which is in line with what Goffman (1981, p. 216) referred to as 'grinding the error into the ground'. Although semantic repair types were also included in the analysis, Plug concluded that the appropriateness *versus* error distinction alone was insufficient to explain the presence or absence of marking (Plug, 2015a, p. 101). Instead, his findings suggest that the prosodic realization of repairs is closely tied to pragmatic, interactional, and discourse-level constraints. This perspective brings a new view that invites to a revision of the static categorical models to include a more dynamic view of self-repair, as socially and contextual phenomenon.

4.2.4 Reparandum Completeness and Prosodic Marking

As discussed in Section 3.2.1, the moment of interruption in repair provides valuable insight into self-monitoring processes. The Dual Loop Theory (Hartsuiker & Kolk, 2001; Levelt, 1989; Levelt et al., 1999) proposes two stages of self-monitoring, one stage in which errors are detected *early* in 'inner speech' (which monitors the speech plan) and, second, one stage that identifies trouble *late,* in 'overt speech' monitoring, which takes place after speech initiation at a post-articulatory phase. Testing what happens at 'inner speech' during mental preparation is a matter of debate because, as noted by Quené and Nooteboom (2024, p. 1) 'out of necessity, this cannot be done without some speculation'. Still, while direct access to inner speech is methodologically limited, researchers have used reparandum completeness as a proxy for it: incomplete reparanda are typically taken to reflect early detection, whereas completed reparanda are associated with late detection (e.g.,Carter & Plug, 2011; Nooteboom, 2010; Nooteboom & Quené, 2017, 2019; Plug, 2016; Plug & Carter, 2014).

Studies using this distinction have reported systematic differences in repair timing and prosody. Plug (2016), for example, found that repairs following incomplete reparanda were produced more rapidly, consistent with early detection and faster response planning (Plug, 2016, p. 533). Similarly, Nooteboom (2010) reported that repairs of interrupted phonological errors were produced with significantly higher pitch and intensity than those of completed ones, suggesting that early repairs are not only faster, but also more likely to be prosodically marked. In contrast, completed repairs tended to be slower, less prosodically salient, and more likely to include editing terms, aligning with detection during overt speech. The analysis of marking based on several measurements of pitch in Hertz (Hz) and intensity in decibels (dB). Delta values were yielded by subtracting pitch and intensity maxima of the first vowel of the repair from those same measurements of the error. A positive delta value for either parameter was interpreted as consistent with marking in repair (recall that Plug & Carter, 2013, followed a similar approach to their intensity and pitch acoustic analyses, and so did Plug, 2014, on pitch, intensity and speech rate).

Nooteboom's (2010) findings were further supported by Nooteboom and Quené (2014) who combined acoustic measurements with perceptual judgements. In this occasion they not only expanded their acoustic analysis but also included perceptual judgements of the reparandum and repair loudness by naïve listeners, providing an additional layer of experimental evidence on how repairs might function to direct listener attention. First, the acoustic analysis included several measurements extracted from the same vowels of both the reparandum and repair; these included: vowel duration (ms), target-to-repair time (ms), maximum and average pitch (Hz) and intensity (dB), and spectral slope (dB) (another dimension of intensity related to energy and vocal effort) (Nooteboom & Quené, 2014, pp. 206-207). Results confirmed previous findings by Nooteboom (2010) on the effects of the detection status (early vs late). The vowels in repairs of early-detected errors have a longer duration, higher maximum intensity and higher average intensity than those of late-detected errors. The spectral slope of the vowels in the repairs was not significantly different from the vowels in the reparandum; however, in early-detected errors, it tends to be less negative as the target-to-repair is longer (Nooteboom & Quené, 2014, p. 210).

Second, the listening experiment asked participants to judge whether the two fragments (i.e., reparandum and repair) were equally loud. Data from 9 listeners was analysed and recoded as 'reparandum louder' and 'repair louder'; response times were also recorded. Results show that again the effects of the detection status were significant on the odds of the repair being judged as subjectively louder than the reparandum (Nooteboom & Quené, 2014, p. 213). Response times were marginally faster for early detected repairs, which was interpreted as these items being subjectively easier for speakers to identify. In sum, Nooteboom and Quené's (2014) results showed that early repairs were perceived as louder and more intense, with longer vowel durations and greater average intensity. Late repairs, on the other hand, were rated as less prominent, and more often included prosodic and lexical cues indicating that an error had been made.

On a more general note, overall findings from Nooteboom's (2010) and Nooteboom and Quené's (2014) studies provide further support for the idea that interrupted (early), repairs are initiated through inner speech monitoring while completed (late) repairs are initiated through overt speech monitoring (Hartsuiker & Kolk, 2001; Levelt, 1983; Levelt et al., 1999). Conversely, these findings do not fit well with Levelt and Cutler's (1983) which proposed that prosodic marking occurs exclusively in lexical repairs, thus excluding the possibility of prosodic marking in phonological repair, a position challenged by Nooteboom's (2010) data. Additionally, Goffman (1981) Plug (2011), Plug (2014) and Plug (2016) also observed that speakers often repair quickly. However, whereas Nooteboom's work focuses on repair initiation timing, these other studies have highlighted repair tempo. Thus, although these two types of 'speed' differ, one concerning onset timing and the other articulation rate, findings across studies converge on the observation that speakers tend to manage errors efficiently, a behaviour observable across both timing dimensions.

4.2.5 Editing Terms and Their Interactional Role in Repair

Editing terms such as *bueno,* in English, 'well', have long been recognized as integral elements of repair sequences, serving not only as placeholders but also as signals to the listener that trouble has been detected. As Levelt (1989, p. 482) describes 'in repair, editing terms play a key role in signalling interlocutors there is trouble'. Several studies have examined the presence, distribution, and timing of these expressions, offering insights into their pragmatic and temporal functions within self-repair.

Work or repair has shown that their use correlates with the timing of error detection. For example, Nooteboom (2010) found that editing terms were more frequently used in repairs of completed errors, cases which, as described earlier, are likely to involve *late* detection during overt speech monitoring In his sample, only 6% of interrupted (i.e., early) repairs included editing terms (7 out of 117), compared to 32% of completed repairs (52 out of 162) (Nooteboom, 2010, pp. 226-227).For example, Nooteboom (2010) found that editing terms were more frequently used in repairs of completed errors, cases which, as described earlier, are likely to involve *late* detection during overt speech monitoring In his sample, only 6% of interrupted (i.e., early) repairs included editing terms (7 out of 117), compared to 32% of completed repairs (52 out of 162) (Nooteboom, 2010, pp. 226-227). Similarly, Nooteboom observed that early repairs tend to be faster, unaccompanied by editing terms, and more likely to be prosodically marked (e.g., via increased pitch and intensity), while late repairs are slower and often include these verbal cues, such as these editing expressions.

Plug (2011) also reported that editing terms tended to be phonetically reduced, reinforcing their low informational status and possibly reflecting a strategy to preserve fluency while gaining extra planning time (Plug, 2011, pp. 293-295). Although his analysis did not directly link editing terms to repair timing, his classification of such elements as informationally redundant reinforces the idea that they play an interactional role rather than a corrective function.

Taken together, these findings support what may be termed a dual strategy hypothesis: when errors are detected early (through inner speech), speakers tend to avoid editing terms and instead use prosodic marking to swiftly resolve the issue (Nooteboom & Quené, 2014, p. 215). In contrast, when detection occurs late (in overt speech), speakers often insert editing terms, potentially as a face-saving device or to signal the need for listener patience as the repair unfolds (cf. Levelt, 1989; Nooteboom, 2010; Plug, 2011). This distinction contributes to the ongoing discussion between the *listener*-oriented and *speaker*-oriented strategies described through the literature. The latter reinforces the view that self-repair is shaped not only by linguistic or semantic factors but also by monitoring mechanisms and the interactional constraints of the moment.

4.2.6 This Study

This chapter aims at contributing to the understanding of prosodic marking in lexical self-repair by studying the phenomena in a new language. Specifically, I explored periodic energy, pitch and speech rate, as acoustic correlates of prosodic marking, in a sample of self-initiated self-repairs extracted from unscripted conversations of Colombian Spanish. The study addresses the following hypotheses:

- **HYPOTHESIS A:** Repairs will more frequently be spoken at a faster tempo than their corresponding reparanda.
- **HYPOTHESIS B:** Repairs of incomplete reparanda will show prosodic marking more frequently than repairs of complete targets.

- **HYPOTHESIS C:** Repairs of complete reparanda are more likely to include editing terms.
- **HYPOTHESIS D:** A semantic classification of repairs, distinguishing between linguistic and factual error repairs versus appropriateness repairs, will not, on its own, predict prosodic marking.

HYPOTHESIS A is in line with findings by Plug (2011, 2014, 2016) on fast tempo (i.e., hypo-speech) in repair and aligns with Goffman's (1981, p. 216) description of 'voiced raised and increased tempo' in strident repairs. The hypothesis addresses Lindblom's (1990, 1996) H&H theory of informativeness on the hypo-hyper continuum, specifically for the case of self-repair. Equally, it is consistent with Plug's definition of marking being a combination between getting the job of correcting done fast and emphasizing the corrected talk (Plug, 2014, p. 13), aligning as well with the above-given description by Goffman's (1981) of strident, or marked, repairs.

HYPOTHESIS B is consistent with Nooteboom's (2010) and Plug's (2016) observations on incomplete (i.e., early detected) repairs to be more frequently prosodically marked than late repairs. It also supports Levelt (1989) and Levelt et al.'s (1999) Dual Loop Perceptual theory and Hartsuiker & Kolk's (2001) Perceptual Loop Model describing two stages of speech monitoring, one that detects troubler early, at inner speech, and one that detects issues late at overt speech; as it is in accordance with the idea that speakers behaviour in relation to prosodic marking is different based on the detection status.

HYPOTHESIS C derives from work by Nooteboom (2010), who proposes that early-detected errors are more likely to be prosodically marked than late-detected errors. Given that late-detected errors were also found by Nooteboom (2010) to be more likely accompanied by editing terms than early-detected errors, I hypothesize that marking is used as an alternative strategy in early-detected trouble to redirect attention to repair without overtly signalling it through an editing term; thus editing terms will more likely be present accompanying unmarked repairs, frequently found in repair sequences of late-detected trouble (i.e., complete reparanda).

HYPOTHESIS D aligns with findings by Plug and Carter (2013) and Plug (2016) who found only weak support for the semantic division between error and appropriateness repairs showing distinct prosodic features in both the reparandum and the repair. In spite of this, I do have some more specific expectations on the patterns that are to emerge in my data; for example, that repairs of errors will probably show marking more frequently than appropriateness repairs (Levelt & Cutler, 1983); also, that linguistic errors will behave differently prosodically from factual errors (Plug, 2016; Plug & Carter, 2013). Still, the general hypothesis is in line with Plug's (2016, p. 53) idea that 'there are specific contexts in which the influence of informativeness [on the semantic dimension, or information salience] is limited by other constraints'; which aligns as well with his (2015a) findings on discourse contexts constraining lexical self-repair and with Goffman's (1981) initial observations on repair prosody.

4.3 Method

This section presents all relevant information regarding the specific methods that were adopted for the analyses of prosodic marking in lexical self-repair. I first describe in detail the nature of the final data set for the analysis of prosody (Section 4.3.1). Subsequently, I present the segmentation and classification of repairs into collections based on semantics, completeness and editing terms (Sections and , respectively). Then, I present all procedures that were followed to, first, extract the speech material from both the reparandum and repair pieces out of the repair sequence; and, second, I present a step-by-sept descriptions of the actions implemented for extracting the necessary phonetic information from each item (4.3.5). Finally, in Section 4.3.5, I present the variables that were incorporated into the statistical analysis, including the relevant choices used for prior class for each parameter that were included in the Bayesian hierarchical models.

4.3.1 Prosodic Marking Data Set

For this thesis I collected unscripted speech from the different dialects and subdialects of Colombian Spanish, based on the dialectal classification proposed by Ruiz Vásquez (2020). Conversational speech was gathered by two interactive tasks, as described in Section . Repairs were sampled from recordings of both tasks as completed by 36 speakers from 5 dialects of Colombian Spanish. For this thesis, 431 instances of lexical self-repair were processed. Before running the analyses, a few repairs were excluded from the sample due to difficulties in their semantic classification, process that I will describe in detail in Section 4.3.3. On some occasions, and due to reparanda being very short, some sequences did not allow for counts of syllables and segments, necessary for the analysis if speech rate and neither allowed for the identification of enough F0 peaks, required for the analysis of the dynamic F0. In consequence, 26 sequences for pitch and periodic energy and 30 sequences for speech rate were removed from the final sample. The rest of the Prosodic Marking data set split into two data frames, one for the analysis pitch and periodic energy comprising 405 items, and one for studying speech rate, which included 401 items.

4.3.2 Segmentation

Each repair was orthographically transcribed by me, the author, a native speaker of Colombian Spanish, then semi-automatically transcribed in Praat (Boersma & Weenink, 2021), using the EasyAlign Spanish extension (Goldman & Schwab, 2014) (See for illustration). Resulting phoneme, word, syllable, and phone tiers were also hand-corrected by me and an additional point tier was added to locate crucial landmarks within the repair sequence, as illustrated in Figure 23. There, the factual error *cuatro* 'four', is corrected to *dos* 'two'; the interval between 1 and 2 is the reparandum; the interval between 4 and 5 is the repair. It is relevant to mention that in the event of having a repair with editing terms, that section of the repair sequence is not included in the acoustic analysis of prosody, and neither are repeated words preceding or following the repair after the resumption of speech.

Figure 22

Resulting five-tiered segmentation of the repair 'Dos pares de zapatos' (Two pairs of shoes).

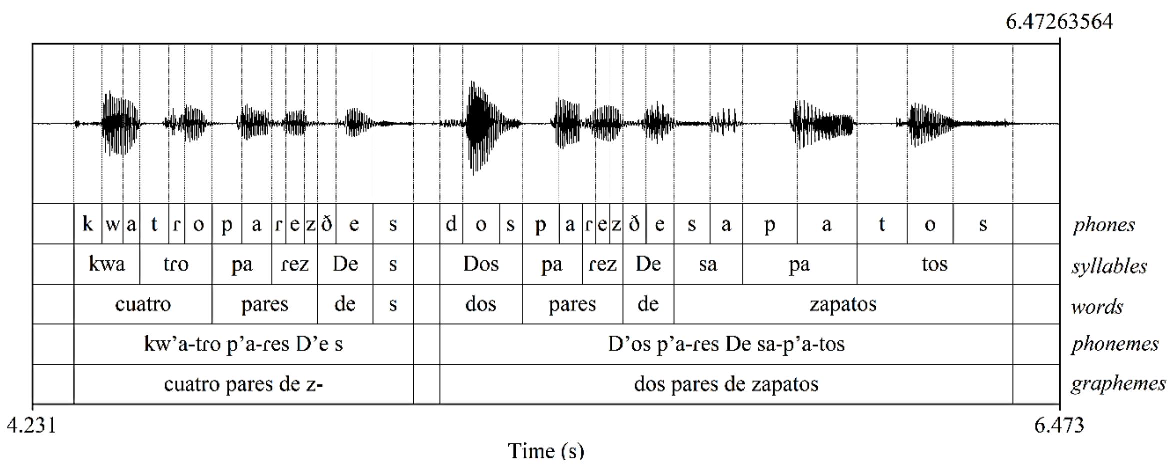
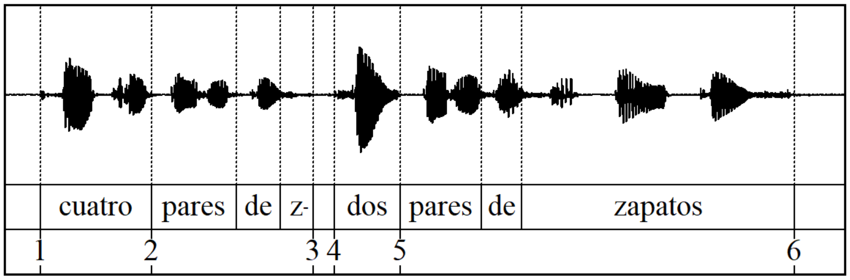


Figure 23

Segmented waveform of the repair 'Dos pares de zapatos' (Two pairs of shoes).



Note. '1' and '2' delimit the reparandum; '4' and '5' delimit the repair; and '6' delimits the end of the repair sequence.

4.3.3 Semantic Classification

The same classification implemented for the analysis of temporal organization, described in Section , was used for the analysis of prosodic marking given that I took the same collection of repairs. Recall that repairs with factual inaccuracies or linguistic ill-formedness are error repairs; in other cases, it was assumed that the first lexical choice is rejected for pragmatic felicity or 'appropriateness' reasons. Also, recall that 104 instances were classified by me, and an independent second rater, a Colombian Spanish discourse analyst. Classifications matched 92 repairs (90%), and consensus was reached for 9 out of the remaining 10; one item was excluded as the raters agreed it could not be reliably classified. Given the high percentage of matching judgements, I made an initial classification of the remaining items, and only doubtful instances were presented to the second rater to reach consensus. In total, 405 items were classified for the analysis of pitch and periodic energy and 401 were classified for the analysis of speech rate. Table 15, below, summarizes the final classification, keeping the subdivision of the full data set into two different data frames, the PITCH AND PERIODIC ENERGY DATA FRAMe, and the SPEECH RATE DATA FRAME.

Table 15

Semantic classification for the analysis of prosodic marking.

		Data Frame	
Repair Type	**Error Type**	**Pitch & Periodic Energy**	**Speech Rate**
Appropriateness		158	157
Error	*Linguistic*	62	62
	Factual	185	182
Total		**405**	**401**

4.3.4 Coding for Editing Terms and Completeness

The same procedure that was used for the analysis of temporal organization was followed in this chapter for coding for *Editing Terms* (as *with* and *without* editing terms) and for *Completeness* (as *complete* vs *incomplete*) (See Section 3.3.4 for details and illustrations on the process). Most instances in the data set did not include editing terms: 285 in the Pitch and Periodic energy data frame and 284 in the Speech rate data frame. The remaining items in both data frames included editing terms: 120 and 117, respectively. For *Completeness*, items in which the first attempt at the target word was stopped before its completion (e.g., the reparandum *señora* ‘lady’, interrupted at the first segment of the second syllable (i.e., [se.ɲ]) and repaired as *dama* ‘dame’, were classified as INCOMPLETE. Conversely, fully articulated reparanda (e.g., *hijo* ‘son’ pronounced in full as [ˈi.xo], and later repaired as *bebé* ‘baby’) were coded as COMPLETE. In total, 231 items in the Pitch and Periodic energy data frame and 229 in the Speech rate data frame were classified as having complete reparanda. The remaining items in both data frames were coded as INCOMPLETE: 174 and 172, respectively. Table 16 and Table 17, below, present a summary of the classification made for both *Completeness* and *Editing Terms*.

Table 16

Final counts of complete vs incomplete repairs included in the analysis of prosodic marking.

	Data Frame	
Completeness	**Pitch & Periodic Energy**	**Speech Rate**
Complete	231	229
Incomplete	174	172
Total	**405**	**401**

Table 17

Final counts of repairs with and without Editing Terms in the analysis of prosodic marking.

	Data Frame	
Editing Terms	**Pitch & Periodic Energy**	**Speech Rate**
Present	120	117
Absent	285	284
Total	**405**	**401**

4.3.5 Selection and Extraction of Phonetic Data

Following segmentation, semantic classification and coding of the repairs, reparanda and repairs excerpts were separated as new audio files, individualizing the relevant pieces of speech. This process allowed for the extraction of the required phonetic data from each item (i.e., pitch, speech rate and periodic energy) thus, further comparisons between the two relevant pieces of speech were possible. In the next paragraphs I will describe in detail the procedure that was followed for processing and extracting data for each acoustic parameter and the reasons that motivated using these specific measures. First, I will explain the procedure for obtaining periodic energy and pitch data using ProPer to later present how speech rate measurements were calculated.

As introduced earlier, the ProPer Workflow (Albert et al., 2024) was employed to gather acoustic data for implementing measurements of periodic energy and pitch into our analysis of repair. One of the motivations for employing ProPer is that it uses periodic energy to enrich the acoustic representation of pitch and prosodic strength. As described earlier, periodic energy is a measurement of the acoustic power of periodic components in the speech signal; this can be thought of as a measurement of general intensity that excludes the contribution of aperiodic noise and transient bursts (Albert et al., 2018, p. 55). The idea of incorporating periodic energy into our analysis is based on Albert's (2023) model proposing that sonority is best understood as a measurement of pitch intelligibility in perception, which is closely linked to periodic energy in acoustics. Albert et al. (2018, p. 805) explain that incorporating periodic energy is motivated by the general auditory perceptual and cognitive principles, according to which periodic energy correlates with pitch intelligibility (See de Cheveigné, 2005; Oxenham, 2012 for details). This view is promising in that descriptions of prosodic weight and prosodic prominence could benefit from this acoustic-based account on sonority, pitch intelligibility and prominence. As described by Albert (2023, p. 142) irrespectively of its relevance to the notion of sonority, the interaction between fundamental frequency (F0) and periodic energy should lead to more comprehensive representations of pitch in speech and beyond.

The ProPer toolbox allows for the incorporation of periodic energy into prosodic research since it extracts periodic energy values, along with other relevant measures, in a quantifiable manner. To my knowledge, this is the first work that has adopted such an approach to the study of prosodic marking in repair. In consequence, the results I report here can contribute to shape future methodologies implementing periodic energy in the acoustic signal not only for the case of repair but also for general prosodic phenomena. Next, I will illustrate how ProPer works[35] and the specific measures that were selected and extracted from it to test our hypotheses.

ProPer implements a workflow that starts by harvesting acoustic data from Praat (Boersma & Weenink, 2021). It uses the information available from Praat's TextGrid segmentations to perform further analysis and produce visualisations in R (R Core Team, 2024). It extracts a wide range of measurements[36] (including periodic energy, mass, relative periodic energy mass, periodic-energy-related F0 measures, relative synchrony and relative ΔF0). For our analyses, I extracted two of periodic-energy-related F0 measurements (i.e., F0 at the Centre of Mass and Dynamic F0, which I will explain in detail next), and periodic energy itself. To obtain those, firstly, files were pre-processed, as recommended, normalizing for loudness at -3.0 dB[37] and adding 200ms[38] of silence on both ends of each audio file and TextGrid to avoid possible issues at analysing file edges in Praat. Next, F0 contours were smoothed using Mausmooth (Cangemi, 2015); this process allowed me to correct pitch points where necessary (e.g., inaccurate identification of voiceless fricatives and trills, octave jumps or phonation types). Mausmooth runs under a Praat script that access all designated audio files, provides

---

[35] I thank Aviad Albert for his kind help in interpreting and analysing ProPer's measurements via personal communication. Any possible errors or misinterpretations in the analysis are solely my responsibility.

[36] See Albert, A., Cangemi, F., Grice, M., & Ellison, T. M. (2024) for details on the different analysis and visualizations that can be extracted from ProPer.

[37] Loudness normalization was performed on Audition (Adobe Inc., 2019)

[38] For an initial subset of repair sequences, 500ms of silence were added at both ends, which did not influence the analysis performed.

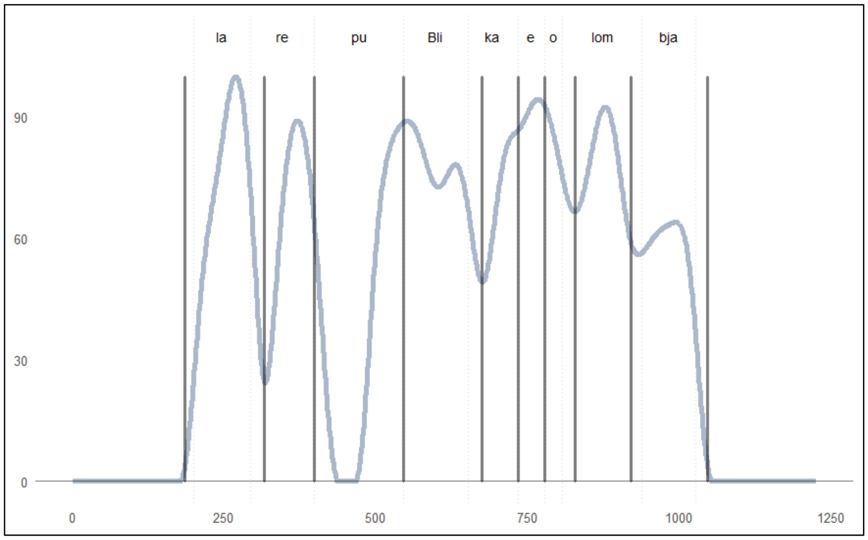
visualizations of the extracted F0, and allows for manual corrections of the pitch contours before proceeding with smoothing, interpolation and the creation of new pitch files. These new files, together with the original TextGrid and audio excerpts for each reparandum and repair were used as input to continue the workflow. In the next step, calculations on the input data are performed to shape the periodic energy curve from the *periodic power vector*[39], which takes place after log-transformation and smoothing. Log-transformation is performed to deal with perception on quantities at various scales and reduce the influence of extreme values. To do so ProPer uses the formula presented in (4), where an optimal periodic floor value for the entire data set is estimated. Smoothing of the resulting periodic energy curve was performed on the 20Hz low-pass filter, to capture segment-size fluctuations. Figure 24 is an example of a periodic energy curve that has been log-transformed and smoothed.

$$periodic\ energy = 10\log_{10}\left(\frac{periodic\ power}{periodic\ floor}\right) \tag{3}$$

[39] Details on the relevant calculations to obtain the *periodic power vector* from the raw intensity of the speech signal are detailed in Albert, A. (2023)

Figure 24

Periodic energy curve of the repair 'República de Colombia' (Republic of Colombia).

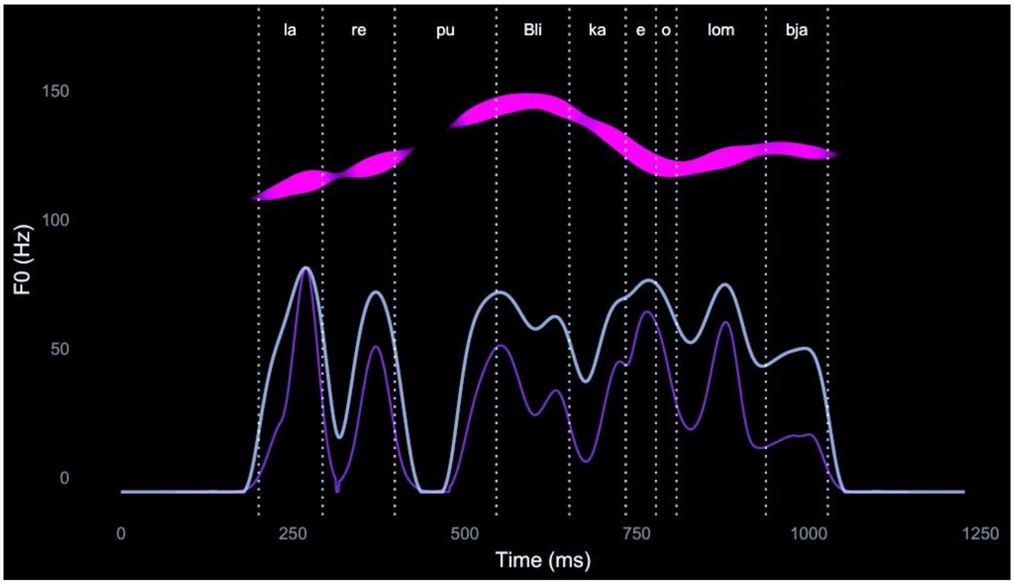


Note. The log-transformed *periodic energy* (light blue) curve is smoothed using a 20 Hz low-pass filter.

The next step was to generate the ProPer visualizations, or PERIOGRAMS, for individual audio files (i.e., reparanda and their corresponding repairs). Periograms (illustrated in Figure 25 are rich visualizations of pitch contours that show the F0 trajectory (where *Frequency* is in the y-axis and *Time* in the x-axis), reflecting the strength of the pitch contour continuously in terms of thickness and darkness of the curve (See Albert et al., 2018, 2019). These periograms also incorporate the periodic energy curve that was also illustrated in Figure 24. At this stage of the workflow, the resulting periograms were used to inspect the periodic energy curves before extracting and saving the quantifiable data, which was used to establish our periodic energy dependent variables.

Figure 25

Periogram of the repair 'República de Colombia' (Republic of Colombia).

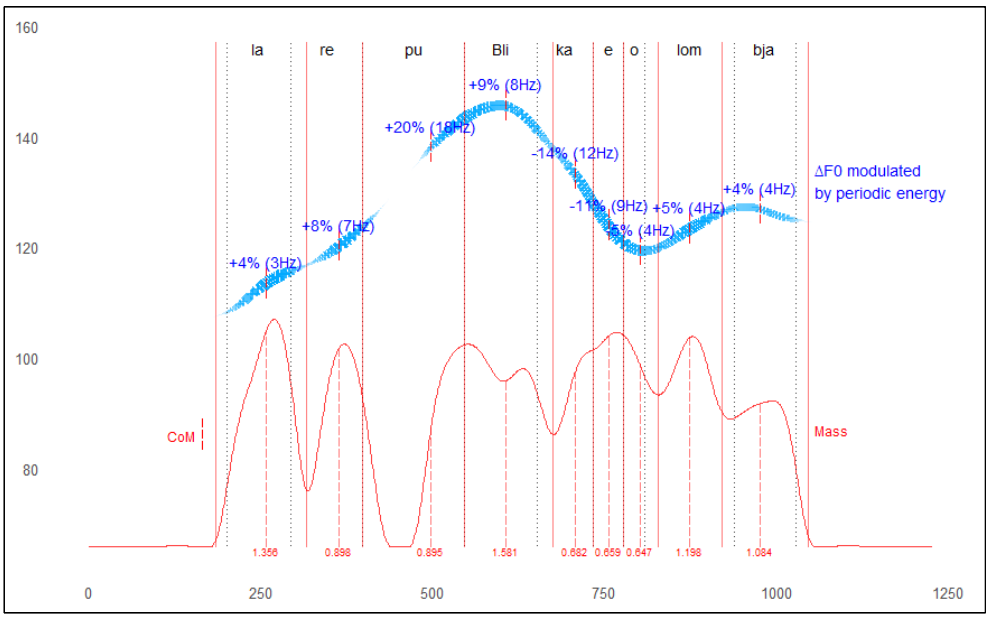


Note. You can observe the *F0 contour* representing the pitch strength (magenta), the *periodic power* (purple), and the log-transformed *periodic energy* smoothed using a 20 Hz low-pass filter (light blue).

In the next step, analyses are performed to extract other relevant information based on the periodic energy curve (e.g., periodic energy mass, synchrony and ΔF0). To our own analyses, periodic-energy-related F0 data is of interest, thus I will briefly describe the measurements that are relevant to obtaining the required F0 information. On the one hand, periodic energy mass (from here on referred to as 'mass') relates to the prosodic strength of each periodic energy fluctuation (i.e., each syllable), quantified as the integral of duration and power. For each fluctuation in the smoothed periodic energy curve, the centre of mass (CoM) is calculated as the average time point weighted by the corresponding periodic energy curve; the CoMs are essential for obtaining other measurements (See Albert, 2023, p. 151 for details). For calculating ΔF0 (Delta F0), for example, ProPer extracts the F0 values at the centres of mass in each utterance and measures the difference in frequency between successive syllables. ΔF0 is obtained as the absolute change (in Hz) and a change relative to each speaker's F0 range (in percentages). Figure 26, below, illustrates the fluctuations from syllable to syllable reflecting the mass, the centres of mass, and the ΔF0 data reflecting the changes in frequency.

Figure 26

Demonstration of ΔF0 *data and Mass in the repair 'República de Colombia' (Republic of Colombia).*



Note. You can observe the ΔF0 data reflecting *change in F0* between syllables (blue). Percentages are given for the relative change, along with absolute values in Hz. *Mass* and *Centre of Mass* (CoM) are also demonstrated (red). Mass values are presented in numbers below each syllabic interval and the dashed vertical red line shows the position of each CoM within intervals.

Having described the extraction of the F0 data from ProPer, for investigating pitch in our repairs, I separated and compared two sets of measurements. First, following common practice, I extracted both *peak values* (minima and maxima) and *average* (mean) F0 at the Centre of Mass (from here on referred to as F0 at CoM) for each reparandum and repair to compute their difference. Second, using ProPer's sequential ΔF0 data, which reflects F0 changes between consecutive syllables, I introduced a new dynamic measurement of pitch variability, applied to both absolute and relative frequency changes. This procedure involved identifying the steepest upward shift in F0 within each reparandum and repair to later compare them. In Figure 26, above, the steepest upward shift corresponds to the syllable [pu] in the word *República* 'Republic', with a 20% relative increase, (18Hz), in comparison to the previous syllable. This implementation follows the proposal by Albert et al. (2018, p. 8.4) to use current tools to make sense of F0 trajectories. With that in mind, capturing and comparing this maximum pitch change can potentially uncover patterns that reflect the prominence that can be associated with resolving issues in self-repair.

Moving on, for speech rate, simple calculations were performed using the information on Praat's TextGrid information. As presented in Section 4.3.2, segmentations included interval tiers for both syllables and segments as well as point tiers marking the start and end of each of reparandum and repair. Counts of syllables and segments as well duration of each reparandum and repair were extracted from Praat's interval tiers and point tiers, respectively. Further calculations were performed in R to obtain speech rates quantified as syllables and segments per second. It is worth noting that ProPer analyses produce their own speech rate measurements. This analysis is performed by calculating the temporal distance between successive CoMs (For details see Albert, 2023, p. 151). Keeping in mind that a few of our reparanda were interrupted and to avoid having items not providing informative measures, we decided to keep the calculations based on smaller units (segments and syllables) rather than extracting speech rates from consecutive CoMs. As a final consideration, keeping in mind Plug et al.'s (2023) findings on listeners' orientation towards canonical forms in speech tempo perception, we opted for using these, which were available from our *syllable* transcriptions (derived from the phonemes tier) rather than using the surface forms on the *phones* tier to perform our speech rate counts (see Figure 22 for reference).

Summing up, we extracted several measurements as our data to analyse each of the selected acoustic parameters associated with prominence (i.e., periodic energy, pitch and speech rate). Once data was gathered, the same procedure was applied to each measurement: comparing the values from each reparandum to those of the corresponding repair by computing the difference between them. This was done by subtracting the reparandum measurements from the repair measurements, yielding a new series of delta values. In this framework, positive deltas are associated with higher prominence where prominence is to be linked with prosodic marking in repair. Table 18 offers a summary of the extracted data and computed deltas.

Table 18

Summary of extracted data for analysing prosodic marking in repair

Acoustic Parameter	**Extracted Data**	**Yielded Deltas**
Periodic energy	Periodic Energy Minimum	ΔPeriodicEnergy_Min
	Periodic Energy Maximum	ΔPeriodicEnergy_Max
	Periodic Energy Mean	ΔPeriodicEnergy_Mean
Pitch	**F0 at CoM (Peak and average)**	
	Minimum F0 at CoM	ΔF0atCoM_Min
	Maximum F0 at CoM	ΔF0atCoM_Max
	Mean F0 at CoM	ΔF0atCoM_Mean
	Dynamic F0	
	Maximum Relative F0 Rise (%)	ΔDynamicF0_Rel
	Maximum Absolute F0 Rise (Hz)	ΔDynamicF0_Abs
Speech rate	Segments per second	ΔSpeechRate_Seg
	Syllables per second	ΔSpeechRate_Syll

4.3.5 Statistical Analysis Using Bayesian Modelling in brms

Similarly to the steps followed in , analyses on periodic energy, pitch and speech rate were performed in R (R Core Team, 2023) by implementing Bayesian regression models using *brms* (Bürkner, 2017) and the *tidyverse* package (Wickham, 2019). As explained in Section , the Bayesian framework was preferred since it offers a series of advantages in comparison to frequentist models. For example, more flexibility in defining models, which allows for complex random effects structures and the possibility of incorporating prior knowledge (i.e., quantitative knowledge from previous studies) into the analysis (Nicenboim & Vasishth, 2016; Vasishth et al., 2018). Again, following Lemoine (2019) and McElreath (2020), priors for all models aimed at weakly informative distributions to regularize results from small sample sizes and incorporate prior knowledge in a conservative way. More details on the advantages of implementing weekly informative priors are given in Chapter 3 (Section 3.3.5). The remainder of this section is structured as follows; first, I will introduce the variables included in the analysis; next, I will describe the model selection process based on data distributions; finally, I will present the priors used for the different model components.

As outlined in Section 4.3.5, a series of measurements (i.e., periodic energy, pitch and speech rate) were selected and extracted from the data to test our prosody-related hypotheses. The derived delta values from each parameter were operationalized as dependent variables; these are: Periodic Energy, Pitch: F0 at CoM and Dynamic F0, and Speech Rate. Once the dependent variables were defined, a series of statistical Bayesian hierarchical models were fitted[40] to test the hypotheses outlined in . All models included *Completeness*, *Editing Terms* and *Repair Type* as predictors (independent variables) and *Speaker* as the only group level effect. Table 19 provides a summary of the variables included in the models.

Table 19

Summary of variables entered into the analysis

Dependent variables	**Periodic Energy, F0 at CoM, Dynamic F0 & Speech Rate**
Independent variables	**Completeness** 1. *Complete* 2. *Incomplete*
	Repair Type 1. *Appropriateness infelicities* 2. *Factual errors* 3. *Linguistic errors*
	Editing Terms 1. *Yes* 2. *No*
Group level effects	**Speaker**

Moving on onto models' selection and fit, in *brms* the argument `family` relates to the distribution of the response variable. Data was inspected and based on their distributions model families were specified. The process was guided by prior and posterior predictive checks. Family selection concurrently led to the models' prior specifications. Most of the data showed heavy-tailed distributions, as was the case for all F0-related data and for the speech rate measures. In

[40] For each dependent variable, several models were fitted to test all sets of calculations extracted (e.g., ΔPeriodicEnergy_Min, ΔF0atCoM_Max, and so on).

those cases, the `family = student()` was chosen given its ability to accommodate outliers. When a normal-like distribution was observed, the `family = gaussian()` was preferred, which was the case for periodic energy minimum and mean[41]. For all tested models, sensitivity analyses[42] comparing alternative model families and priors' specifications indicated that the chosen distributions consistently provided a better fit. Table 20 outlines each model's prior class and the choices made for every parameter.

[41] Periodic energy maxima distribution showed heavy tails; therefore, extreme outliers were filtered and removed, and a student-t family was used.

[42] Available in the OFS repository.

Table 20

Family and priors set for the models evaluating the effects of predictor variables on Periodic Energy, F0 at CoM, Dynamic F0 and Speech Rate

Variable	Family, Prior Class and Settings

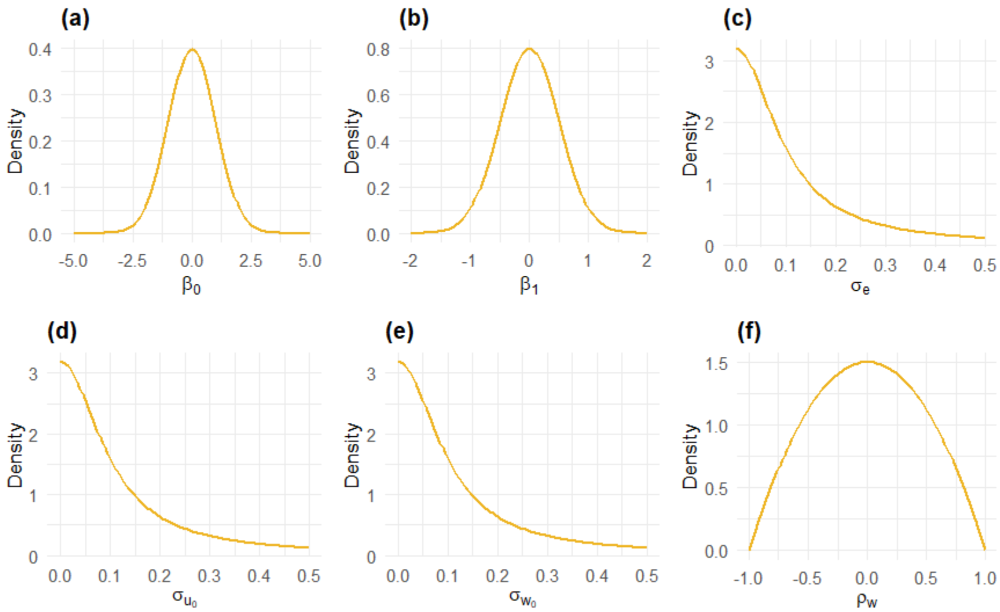
Periodic Energy
ΔPeriodicEnergy_Min
ΔPeriodicEnergy_Mean
	`family = gaussian()`
	
	`priors <- c(`
	`prior(normal(0, 0.5), class = b),`
	`prior(normal(0, 1), class = Intercept),`
	`prior(cauchy(0, 0.1), class = sigma),`
	`prior(cauchy(0, 0.1), class = sd),`
	`prior(lkj(2), class = cor))`

Periodic Energy
ΔPeriodicEnergy_Max
	`family = student()`
	
	`priors <- c(`
	`prior(normal(0, 0.5), class = b),`
	`prior(normal(0, 1), class = Intercept),`
	`prior(cauchy(0, 0.1), class = sigma),`
	`prior(cauchy(0, 0.1), class = sd),`
	`prior(gamma(2, 0.5), class = nu),`
	`prior(lkj(2), class = cor))`

F0 at CoM
ΔF0atCoM_Min
ΔF0atCoM_Max
ΔF0atCoM_Mean
Dynamic F0
ΔDynamicF0_Rel
ΔDynamicF0_Abs
Speech Rate
ΔSpeechRate_Seg
ΔSpeechRate_Seg
	`family = student()`
	
	`priors <- c(`
	`prior(normal(0, 0.5), class = b),`
	`prior(normal(0, 5), class = Intercept),`
	`prior(cauchy(0, 0.25), class = sigma),`
	`prior(cauchy(0, 0.1), class = sd),`
	`prior(gamma(2, 0.5), class = nu),`
	`prior(lkj(2), class = cor))`

Figure 27

Prior distributions for the parameters in the linear mixed models for Periodic Energy (Gaussian).

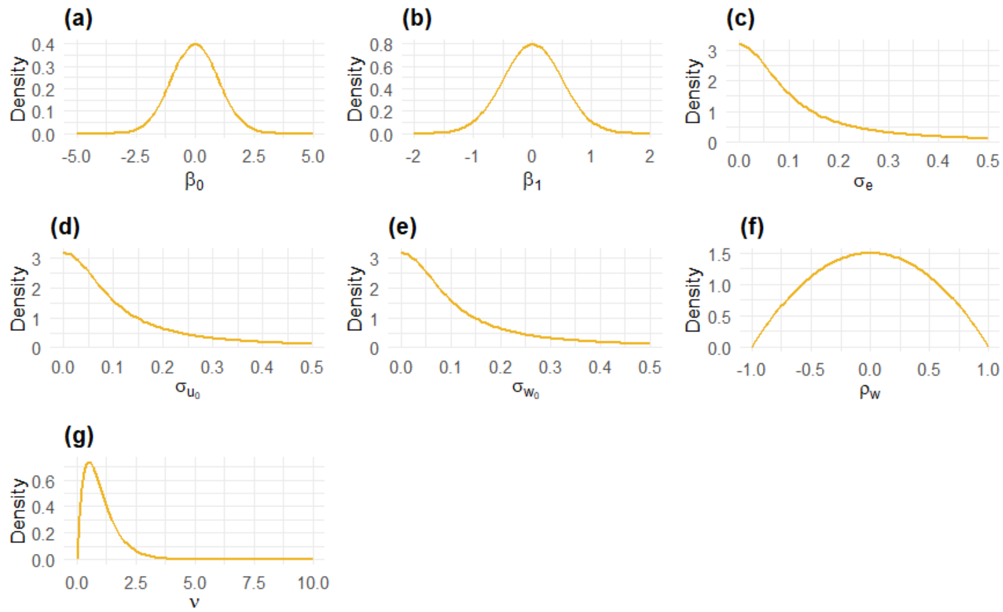


Note. The prior for the grand mean parameter $\boldsymbol{\beta_0}$ (a) follows a *Normal* (0, 0.5) distribution; the prior for the parameter representing the effect of the intercepts $\boldsymbol{\beta_1}$ (b) follows a *Normal* (0, 0.5). The priors of the standard deviation $\boldsymbol{\sigma_e}$ (c) and for the standard deviations of the group-level effects $\boldsymbol{\sigma_u}$ (d) and $\boldsymbol{\sigma_{wo}}$ (e) are all *Cauchy* (0, 0.1). Finally, the prior for the correlation between the random effects $\boldsymbol{\rho_w}$ (f) follows a LKJ (v = 2) distribution.

I aimed at establishing weakly informative priors to allow regularization without constraining the models. This was particularly important given that the analysis of prosodic marking covered varied measurement ranges. The latter differs from the report presented on the temporal organization of repair (Chapter 3) where both dependent variables dealt with similar timing intervals. Priors chosen for the analysis of all variables included in the analysis of prosodic marking can be visualized in Figure 27, Figure 28 and Figure 29 below.

Figure 28

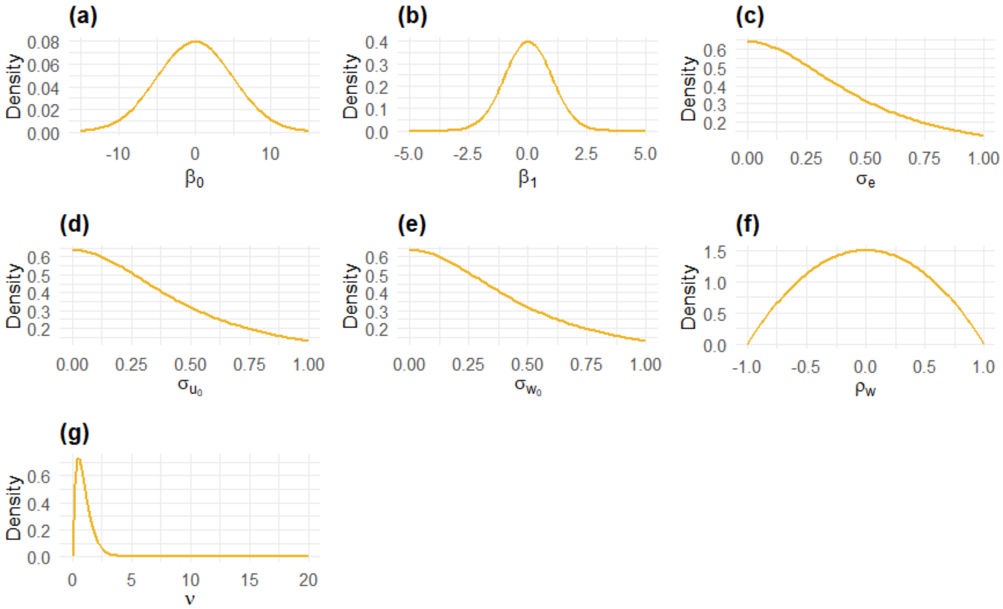
Prior distributions for the parameters in the linear mixed models for Periodic Energy (Student-t).



Note. A *Gamma* (2, 0.5) prior was introduced for the degrees of freedom parameter **ν** (g), while the rest of the priors remained the same as in Figure 27.

Figure 29

Prior distributions for the parameters in the linear mixed models for Speech Rate and Pitch (Student-t).



Note. The prior for the grand mean parameter $\boldsymbol{\beta_0}$ (a) follows a *Normal* (0,5) distribution; the prior for the parameter representing the effect of the intercepts $\boldsymbol{\beta_1}$ (b) follows a *Normal* (0, 0.5). The priors of the standard deviation $\boldsymbol{\sigma_e}$ (c) follow *Cauchy* (0, 0.25). The distributions of the standard deviations of the group-level effects $\boldsymbol{\sigma_u}$ (d) and $\boldsymbol{\sigma_{wo}}$ (e) are all *Cauchy* (0, 0.1). The prior for the correlation between the random effects $\boldsymbol{\rho_w}$ (f) follows a LKJ (v = 2) distribution. Finally, a *Gamma* (2, 0.5) prior was introduced for the degrees of freedom parameter **ν** (g).

The class `Intercept` (i.e., β_0) are the intercept parameters (i.e., dependent variables). The priors for the intercepts were set as *Normal*, varying according to the measurements' ranges. In the case of periodic energy, both for the `gaussian` and `student-t` families, the prior was set as (*Normal* (0, 1) aiming at a lower intercept for the obtained bounded scale (Figures and , all other models used a (*Normal* (0, 5) prior (Figure 29). The class `b` (i.e., β_1) refers to the slopes in the model; that is, the slopes for the effects of interest (*Completeness*, *Repair Type* and *Editing Terms*). This prior was also regularized for all models so that results of the effects on the different dependent variables are comparable; the prior was set a normal distribution centred at zero with SD = 0.50 (*Normal* (0, 0.5) (Figures 27 and 29); which is a wide prior that allows for an ample range of differences between the conditions (Winter & Bürkner, 2021, p. 15). The standard deviation, class `sigma` (σ_e) was given non-informative priors, again, following Lemoine (2019, p. 918). I set a *Cauchy* distribution centred at zero with a SD=0,25 (*Cauchy* (0, 0.25) (, for the measurements of pitch and speech rate, and a slightly more constrained choice was made for periodic energy (*Cauchy* (0, 0.1) (Figures 27 and 28). Similarly, the parameter class `sd`, the standard deviation for the group level effects (i.e., σ_u and σ_{wo}), was placed at *Cauchy* (0, 0.1) for all models (Figures , and), following Lemoine's (2019, p. 925) description on the potential of the Cauchy distributions for yielding proper posterior estimates across different sample sizes. For the correlation parameter (ρ_w), class `corr`, I went for the standard choice LKJ(2) (Figures 27, and 29) assuming that extreme values are unlikely (Vasishth et al., 2018, p. 150). Finally, for models following a `student-t` distribution, a *Gamma* (2, 0.5) prior was specified for the degrees of freedom (i.e., ν) of the parameter class = `nu` (Figures 28 and 29). This choice aimed at controlling the heaviness of the tails in the distribution, balancing robustness against outliers while maintaining flexibility in modeling variability (Gelman et al., 2013).

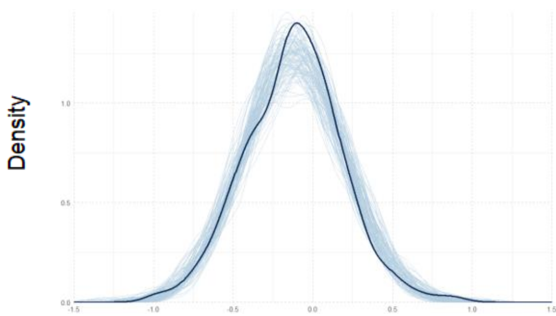
Following fit, all models were estimated via Markov Chain Monte Carlo (MCMC). Four sampling chains were run for 8000 iterations, with a warm-up period of 4000 iterations. Adapt-delta was increased in the `control` argument to avoid any divergent transitions. In the final models all chains mixed well (Rhat = 1.0). Posterior predictive checks, which assess how well the model matches up with the observed data (Vasishth et al., 2018), show that the predicted and observed data for all models have similar distributions (see Figure 30). Therefore, it can be concluded that the models have a reasonable fit.

Figure 30

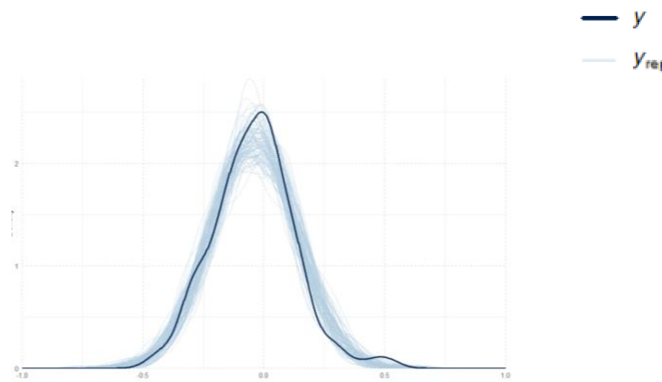
Posterior predictive checks for the modelled data.

Periodic Energy

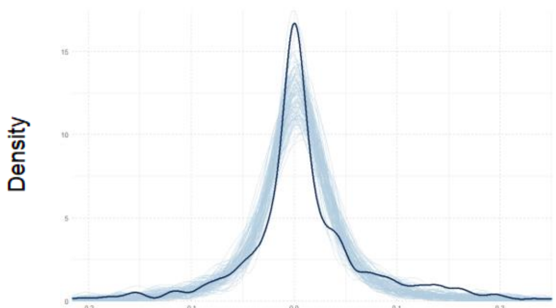
ΔPeriodicEnergy_Min



ΔPeriodicEnergy_Mean

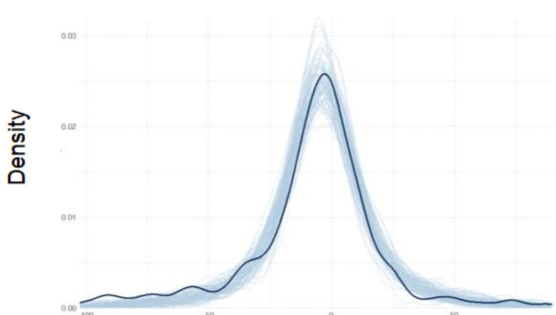


ΔPeriodicEnergy_Max

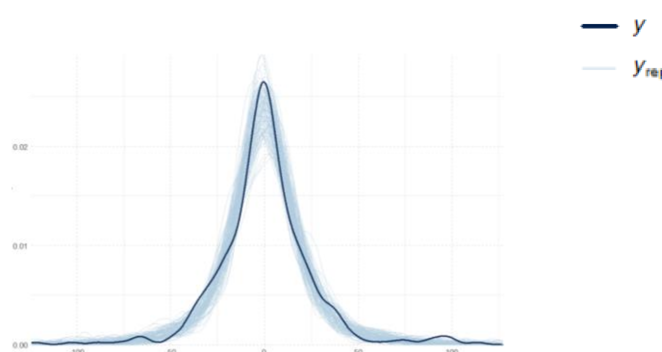


F0 at CoM

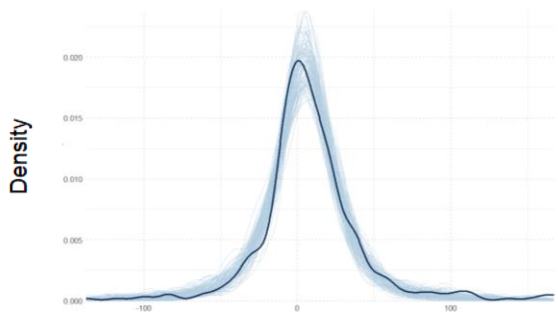
ΔF0atCoM_Min



ΔF0atCoM_Mean

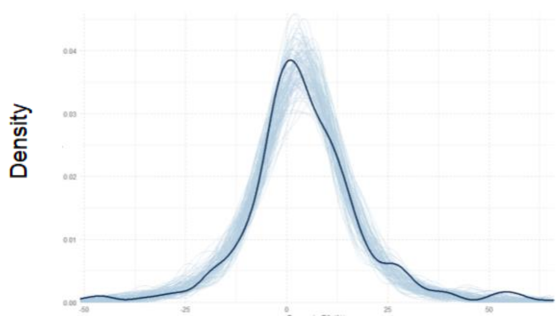


ΔF0atCoM_Max

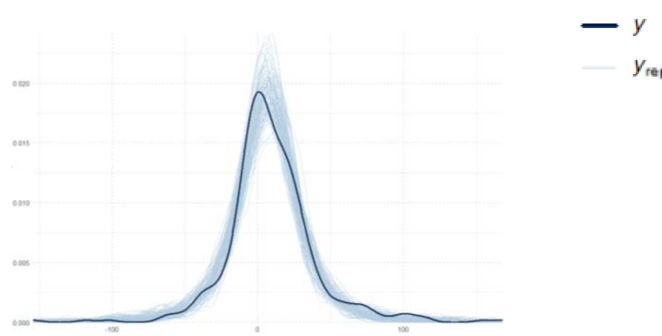


Dynamic F0

ΔDynamicF0_Rel

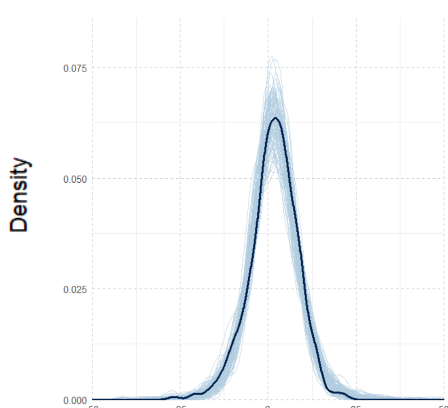


ΔDynamicF0_Abs

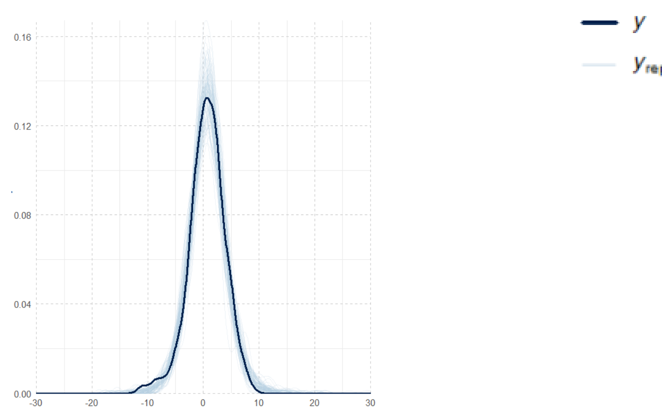


Speech Rate

ΔSpeechRate_Seg



ΔSpeechRate_Syll



Note. The lines marked y_{rep} refer to the posterior predictive values generated by the models, and the black solid line are the observed data.

4.4 Results

Pearson correlations tests were conducted to assess the relationships between the individual Deltas yielded extracted from each dependent variable (i.e., Periodic Energy, Pitch and Speech Rate).

On PERIODIC ENERGY, results revealed that all three measurements (i.e., ΔPeriodicEnergy_Min, ΔPeriodicEnergy_Mean and ΔPeriodicEnergy_Min and ΔPeriodicEnergy_Max) are significantly correlated, although not at the same level. *Minimum* and *mean* periodic energy had a strong positive correlation (r=0.883, $p < 0.001$) indicating that lower periodic energy values tend to co-vary with their mean; the relationship between *mean* and *maximum* periodic energy was moderately positive (r=0.499, $p < 0.001$), suggesting that mean values also tend to increase with higher peaks of periodic energy. Although *minimum* and *maximum* were weakly correlated, the relationship was still significant (r=0.156, $p = 0.002$). For the measurements of PITCH, I tested for correlations in the measurements extracted for F0 at CoM (i.e., ΔF0atCoM_Min, ΔF0atCoM_Mean and ΔF0atCoM_Max) and for Dynamic F0 (i.e., ΔDynamicF0_Rel, in percentages; and, ΔDynamicF0_Abs, in Hertz. For F0 at CoM, *mean* had a very strong correlation with both *maximum* ($r = 0.839$, $p < 0.001$) and *minimum* ($r = 0.821$, $p < 0.001$) indicating that the mean F0 at CoM is heavily influenced by its minimum and maximum values. *Minimum* and *maximum* showed a moderate positive correlation ($r = 0.454$, $p < 0.001$) which shows that as minimum values increase, maximum values tend to rise as well. For Dynamic F0, the absolute and relative measurements, as expected, showed a highly strong correlation ($r = 0.934$, $p < 0.001$). Lastly, on SPEECH RATE, two Deltas were tested: ΔSpeechRate_Syll (Syllables per second) and ΔSpeechRate_Seg (Segments per second). The results revealed a strong positive correlation between the two of them ($r = 0.823$, $p < 0.001$), indicating that as the number of syllables per second increases, the number of segments per second also tends to increase.

Overall, the significant correlations found across periodic energy, pitch, and speech rate measurements suggest that these acoustic features are systematically interrelated. Particularly, speech rate and periodic energy measures appear to capture similar properties of speech *speed* and *strength*, respectively. Furthermore, the strong associations found within F0 at CoM and Dynamic F0 highlight how different pitch measures reflect the same underlying prosodic patterns. The patterns found contribute to our understanding of how acoustic parameters can interact in prosodic analysis and reinforce their validity as interpretable indicators of prosodic variation. Table 21 offers a summary of the of the Pearson's correlation tests performed, including significance and confidence intervals, and Appendix D presents a Scatterplot Matrix for each group of variables compared.

Table 21

Summary of the relationships found between acoustic measurements.

	Relationship	***r***	**CI 95%**	**p-value**
Periodic Energy	ΔPeriodicEnergy_Min ΔPeriodicEnergy_Mean	0.883	[0.860, 0.903]	$p < 0.001$
	ΔPeriodicEnergy_Min ΔPeriodicEnergy_Max	0.156	[0.059, 0.249]	$p = 0.002$
	ΔPeriodicEnergy_Mean ΔPeriodicEnergy_Max	0.499	[0.422, 0.569]	$p < 0.001$
F0 at CoM	ΔF0atCoM_Min ΔF0atCoM_Mean	0.821	[0.786, 0.851]	$p < 0.001$
	ΔF0atCoM_Max ΔF0atCoM_Mean	0.839	[0.807, 0.865]	$p < 0.001$
	ΔF0atCoM_Min ΔF0atCoM_Max	0.454	[0.373, 0.528]	$p < 0.001$
Dynamic F0	ΔDynamicF0_Abs (Hz) ΔDynamicF0_Rel (%)	0.934	[0.920, 0.945]	$p < 0.001$
Speech Rate	ΔSpeechRate_Syll ΔSpeechRate_Seg	0.823	[0.789, 0.853]	$p < 0.001$

Moving on onto the results from our exploratory Bayesian regression models, I will now present a brief summary of the general observations made. As expected from the observed significant positive correlations, for Speech Rate, the related models that measured the effect of the independent variables on ΔSpeechRate_Syll (syllables per second) and ΔSpeechRate_Seg (segments per second) showed similar coefficients and effects, pointing to analogous results. The same was the case for the effects of the predictors on Dynamic F0 (ΔDynamicF0_Rel and ΔDynamicF0_Abs), as both models showed equal tendencies in both their overall prediction and the effects of the independent variables. On F0 at CoM, the effects of the predictor variables showed equal tendencies for all three measurements (i.e., ΔF0atCoM_Min, ΔF0atCoM_Mean and ΔF0atCoM_Mean); however, the overall predictions of the model changed when we moved from the *minimum* to the *maximum* values. Finally, results on Periodic Energy showed that when we measured *minimum* (ΔPeriodicEnergy_Min) and mean (ΔPeriodicEnergy_Min) values, the overall tendencies and effects of the predictors followed a similar trend with incomplete items showing a lower periodic energy. Also, on the semantic distinction of repairs, factual and error repairs exhibited higher energy values than appropriateness repairs. These findings align with the strong correlation found between them; however, there was a contrast in the robustness of the results on semantic distinction in minimum and mean values. Even though both models capture the same trend; we cannot confirm that the effect of the semantic variable on the mean are robust since confidence intervals cross zero. Conversely, results obtained from the *maximum* values (ΔPeriodicEnergy_Max) showed no effects with most of the effects exhibiting values close to zero.

Based on the above findings, I will report the main results of this chapter using the most relevant and representative models for each prosodic parameter. On PERIODIC ENERGY, given that

ΔPeriodicEnergy_Min and mean ΔPeriodicEnergy_Min exhibited similar trends, I will focus on the results obtained from the *mean* values. The reason for this is that we are analysing lexical repairs, in which reparandum and repair can have a different phonotactic structure across syllables, therefore, using the average measure provides a more reliable contrast than the peak values. Nevertheless, I will discuss the implications of the patterns observed in the minimum and maximum periodic energy as well.

To report on PITCH, I will present results from the model using the absolute measure of F0 change in Hertz (ΔDynamicF0_Abs) for Dynamic F0. For F0 at CoM, I will focus on the findings obtained from the models that observed the mean (ΔF0atCoM_Mean), as they align closely with the dynamic F0 measurements. For SPEECH RATE, I will focus on the model analysing ΔSpeechRate_Seg (segments per second). Full summaries of all tested models are available in the OFS repository accompanying this project and also included on

Having selected the relevant measurements for the following section of this report, I will present their distribution along with the mean estimates derived from one-sample t-tests. These tests were performed to determine whether the selected prosodic measurements exhibited systematic changes. The conducted one-sample t-tests compared the mean of each measurement to zero. The results revealed that ΔPeriodicEnergy_Mean ($M = -0.042$, $p < 0.001$) showed a statistically significant negative shift, indicating a systematic decrease in periodic energy. Similarly, ΔDynamicF0_Abs (Hz) ($M = 9.60$, $p < 0.001$) and ΔSpeechRate_Seg (Segments per second) ($M = 1.57$, $p < 0.001$) both showed significant positive shifts, suggesting a systematic increase in these prosodic features. In contrast, ΔF0atCoM_Mean ($M = 0.73$, $p = 0.58$) was the only measurement that did not significantly differ from zero, indicating that pitch variation around the Centre of Mass (CoM) was not systematically changed in the data. These findings highlight the systematic prosodic adjustments made by speakers in the context of our repair sequences, particularly in dynamic F0, Periodic Energy, and Speech Rate. Figure 31, Figure 32., and Figure 34 below, show the distribution of each tested measurement.

Figure 31

Distribution of ΔPeriodicEnergy_Mean. The red dashed line represents zero.

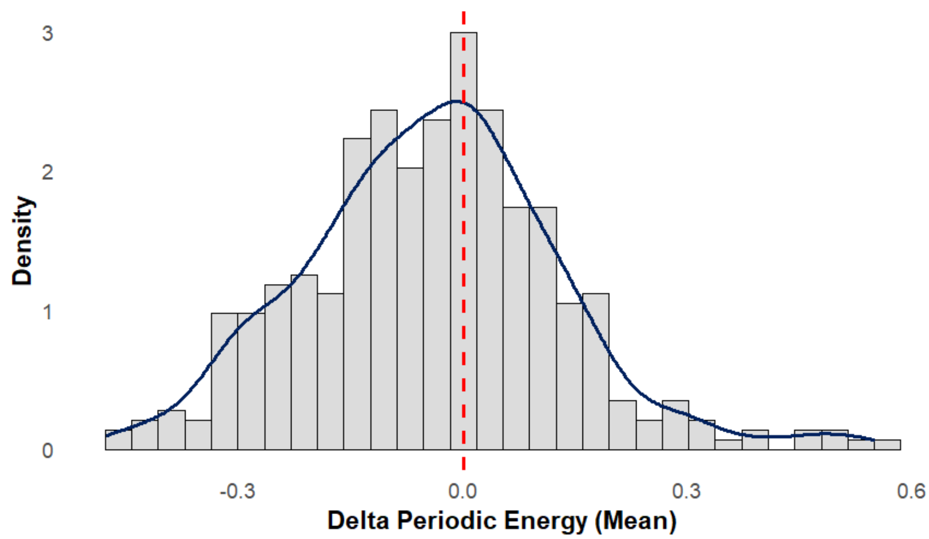


Figure 32

Distribution of ΔF0atCoM_Mean. The red dashed line represents zero.

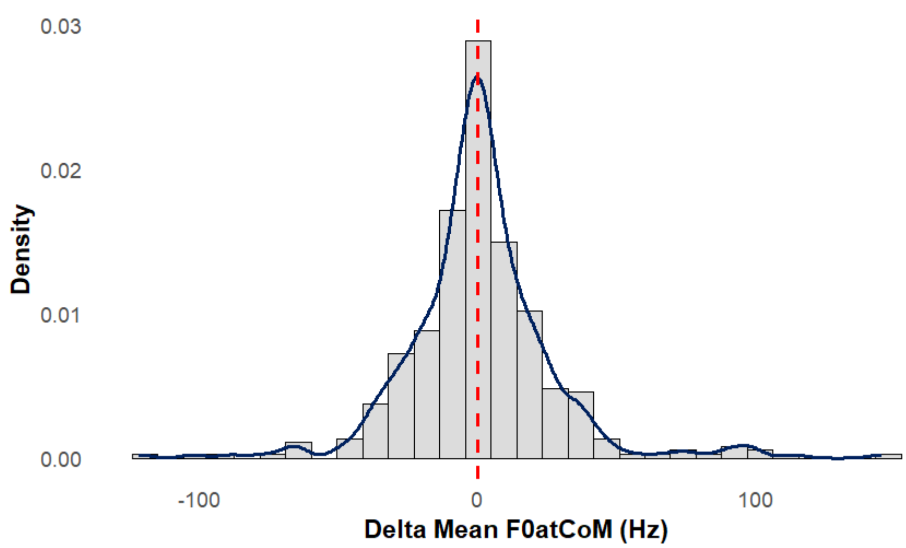


Figure 33

Distribution of ΔDynamicF0_Abs. The red dashed line represents zero.

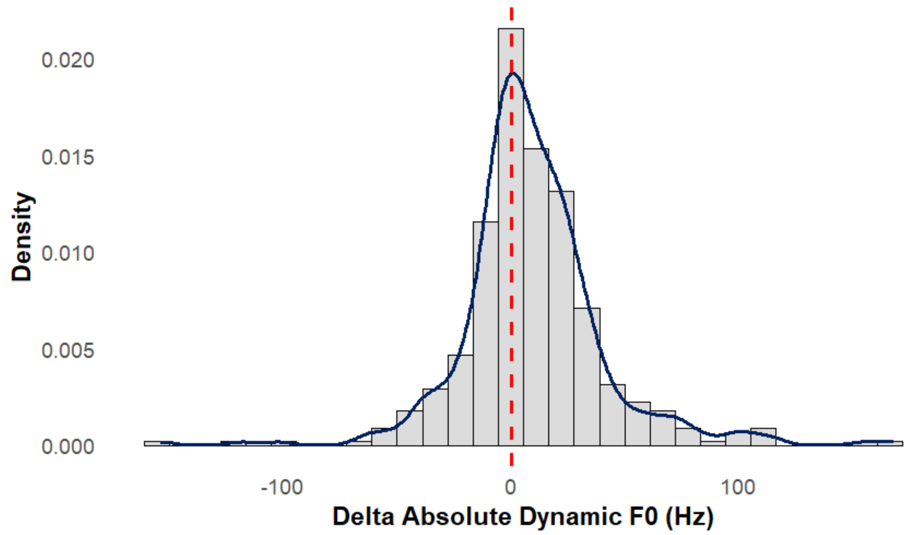
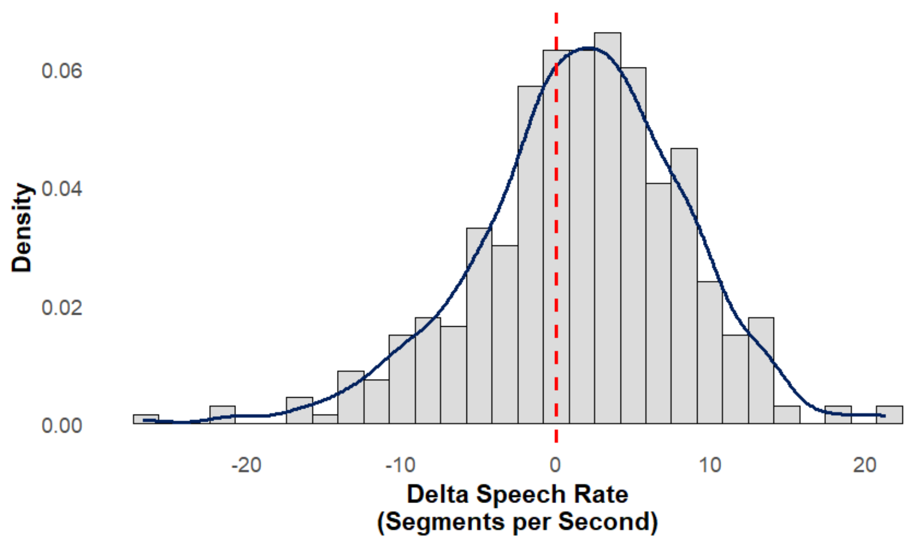


Figure 34

Distribution of ΔPeriodicEnergy_Mean. The red dashed line represents zero.

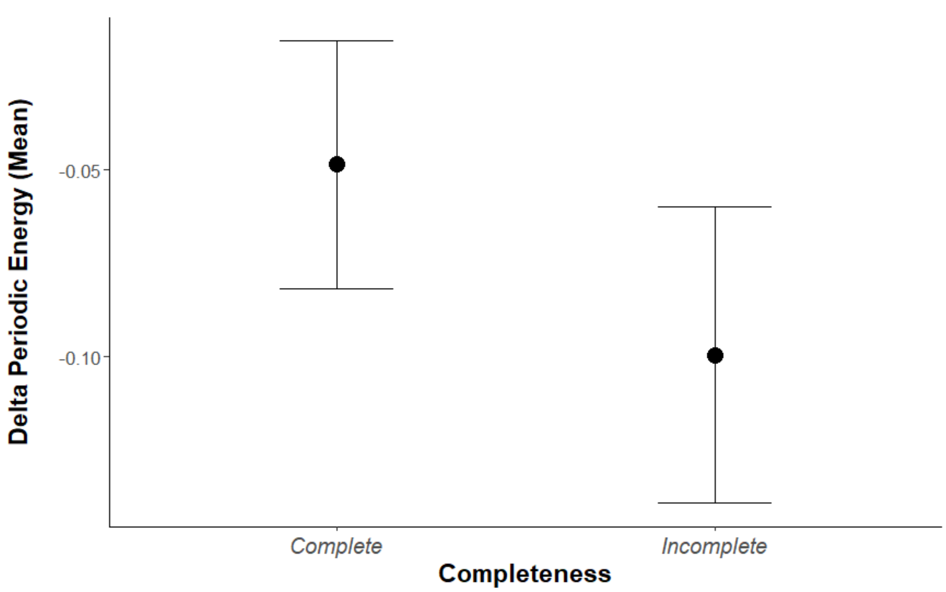


4.4.1 Relevance of Completeness, Editing Terms and Repair Type on Periodic Energy

After evaluating the effects of all independent variables on Periodic Energy, results show that the model evaluating the mean values of Periodic Energy (ΔPeriodicEnergy_Mean) indicated robustness on the effect of *Completeness*, showing a reduction on the strength of repairs in respect to reparanda, when the these were incomplete with an estimate of -0.05 (*SE* = 0.02, 95% CI [-0.09, -0.01]). The 95% credible interval does not include zero, suggesting a robust impact of this variable on Periodic Energy, as the effect is unlikely to be zero. Given that our periodic energy delta is a measure of sonority and syllabic strength yielded from a bounded 0-1 scale, this result implies that repairs with incomplete reparanda tend to be produced with weaker syllabic prominence (less prosodic strength) compared to their complete counterparts. Figure 35 illustrates the effect of Completeness on Periodic Energy as extracted from the model analysing mean values.

Figure 35

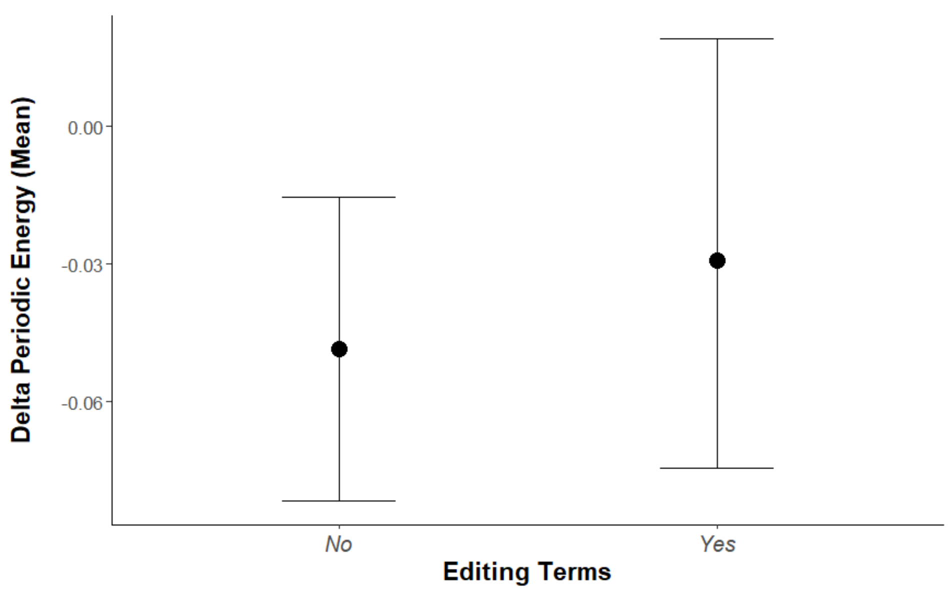
Conditional effects plot for Completeness as extracted from the Gaussian model analysing ΔPeriodicEnergy_Mean. The error bars display 95% credible intervals; the dots represent posterior means.



Moving on to *Editing Terms*, analysis reveals no major difference on the periodic energy levels between the reparandum and repair based on the presence or absence of editing terms. Repair sequences that included editing expressions showed a small increase on periodic energy of approximately 0.02 (*SE* = 0.02 and a CI: [-0.02, 0.7]) compared to sequences with no editing expressions. However, since the credible interval includes a zero, this suggests that the effect is not robust, indicating that the observed increase in the repairs' prosodic strength when editing terms are present is negligible. Figure 36 illustrates the impact of *Editing Terms* on periodic energy mean.

Figure 36

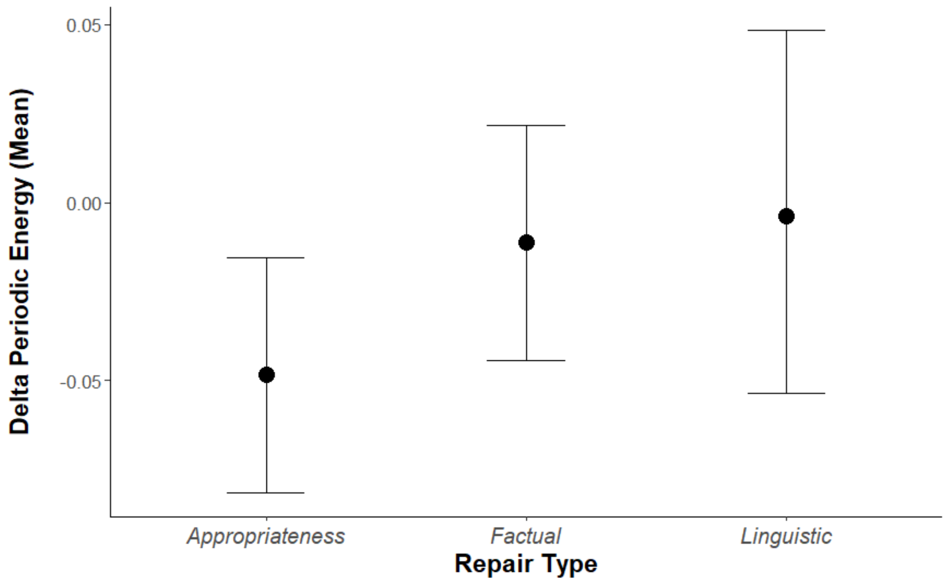
Conditional effects plot for Editing Terms as extracted from the Gaussian model analysing ΔPeriodicEnergy_Mean. The error bars display 95% credible intervals; the dots represent posterior means.



Regarding *Repair Type*, appropriateness repairs showed a decrease in periodic energy, with a mean estimate of approximately -0.05 (SE = 0.02, 95% CI [-0.08, -0.02]). In contrast, both factual and linguistic repairs exhibited increased values, suggesting they may be prosodically stronger than appropriateness repairs. Factual repairs had a mean estimate of 0.04 (*SE* = 0.02, 95% CI [0.00, 0.08]), while linguistic repairs showed a slightly higher estimate of 0.05 (*SE* = 0.03, 95% CI [-0.01, 0.10]). However, since the credible intervals for both factual and linguistic repairs include zero, the evidence for a reliable difference remains inconclusive. In this model, appropriateness repairs are the reference category (intercept), and the estimates for factual and linguistic repairs are expressed as deviations from that baseline. Thus, the posterior means for factual and linguistic repairs, computed as intercept plus the effect, are approximately -0.01 and 0.00, respectively. Figure 37 illustrates the impact of the different semantic repair subtypes on ΔPeriodicEnergy_Mean.

Figure 37

Conditional effects plot for Repair Type as extracted from the Gaussian model analysing ΔPeriodicEnergy_Min. The error bars display 95% credible intervals; the dots represent posterior means.



Once the main analysis was performed and given the robust effects of *Completeness* and the trend observed for *Repair Type*, the model on periodic energy mean was further refined to test potential relationships between predictors. Despite exploring interactions, no strong evidence of a robust relationship between *Completeness, Editing Terms* or *Repair Type* was found. For that reason, the model analyzing ΔPeriodicEnergy_Mean was kept simple with no further interaction effects reported[43].

[43] The refined models on ΔPeriodicEnergy_Mean, which explored interactions, and their outcomes are found in the corresponding R markdown in the OFS repository.

To sum up, a complete outline of estimates for the variables *Completeness*, *Editing Terms* and *Repair Type* on ΔPeriodicEnergy_Mean is given in Table 22, which includes a brief interpretation of the findings.

Table 22

Posterior mean, standard error, 95% credible interval and brief interpretation for each predictor in the model analysing ΔPeriodicEnergy_Mean.

	Estimate	*SE*	Lower bound	Upper bound	Interpretation
Intercept	-0.05	0.02	-0.08	-0.02	*Baseline periodic energy for complete repairs, appropriateness repairs, without editing terms.*
Completeness Incomplete	-0.05	0.02	-0.09	-0.01	*Repairs with* incomplete *reparanda have a lower periodic energy by 0.05 units compared to complete ones.*
Editing Terms Yes	0.02	0.02	-0.02	0.07	*Potential increase of 0.02 units in periodic energy with* editing terms*, but with uncertainty.*
Repair Type Factual	0.04	0.02	-0.00	0.08	*Factual errors increase periodic energy by 0.04 units, indicating higher prosodic strength than in appropriateness repairs but with uncertainty.*
Repair Type Linguistic	0.05	0.03	-0.01	0.10	*Linguistic errors increase periodic energy by 0.11 units, making them prosodically stronger than appropriateness repairs but with uncertainty.*

As explained earlier, minimum and mean values of periodic energy showed robust effects for *Completeness* in both models and for *Repair Type* in periodic energy minimum. Conversely, the model that tested the effects of our predictors on periodic energy maximum exhibited negligible results when reparanda and repairs were contrasted in terms of their prosodic strength.

Tight credible intervals that include zeros suggests that these effects could range from small to no change, meaning predictors barely influence *maximum* periodic energy. See Table 23 for a full summary on the predictor coefficients on the model evaluating ΔPeriodicEnergy_Max.

Table 23

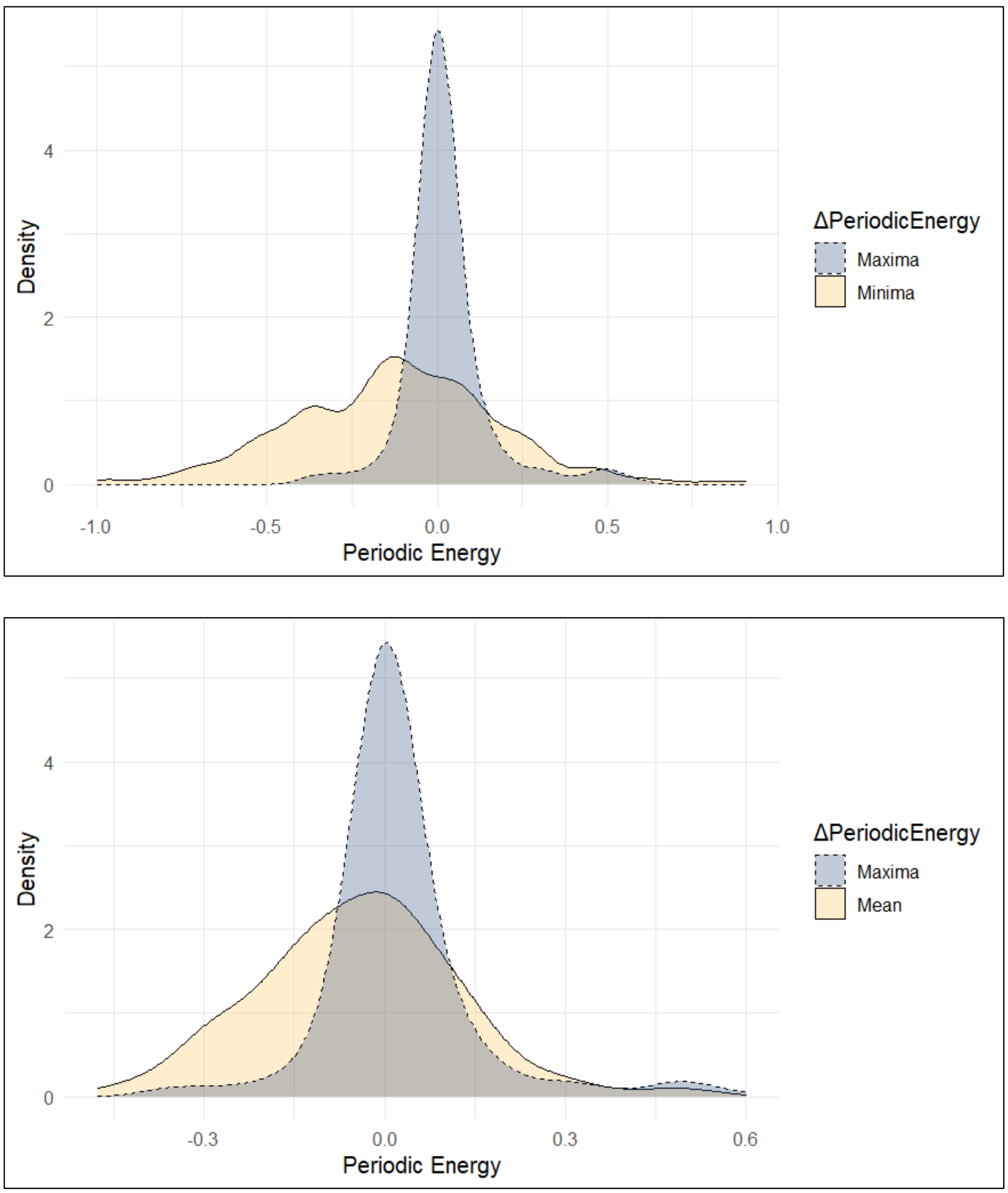
Posterior mean, standard error, 95% credible interval and brief interpretation for each predictor in the model analysing ΔPeriodicEnergy_Max.

	Estimate	***SE***	**Lower bound**	**Upper bound**	**Interpretation**
Intercept	0.00	0.00	-0.00	0.01	*Baseline periodic energy for complete repairs, appropriateness repairs, without editing terms.*
Completeness Incomplete	-0.00	0.00	-0.01	0.00	*Repairs with* incomplete *reparanda show negligible difference in periodic energy compared to complete ones.*
Editing Terms Yes	0.00	0.00	-0.01	0.01	*No significant impact of* editing terms *on periodic energy.*
Repair Type Factual	0.00	0.00	-0.01	0.01	*Factual errors do not significantly affect periodic energy compared to appropriateness repairs.*
Repair Type Linguistic	0.01	0.01	-0.00	0.02	*A minor potential increase of 0.01 units in periodic energy for* linguistic repairs*, but with uncertainty.*

The contrast in the results from models on *minima* and *mean* and the results from *maxima* are in line with the weak correlation found earlier between ΔPeriodicEnergy_Min and ΔPeriodicEnergy_Max (Pearson's r = 0.156); which contrasted with the strong positive correlation between mean and minima (Pearson's r = 0.883). This appears to indicate that results on the maxima measure are shaped by separate underlying conditions. This also becomes apparent when distributions are observed together. In Figure 38, which contrasts *maxima* to both *mean* and *minima* distributions, we can evidence that while ΔPeriodicEnergy_Min and ΔPeriodicEnergy_Mean values are broadly spread, ΔPeriodicEnergy_Max values are clustered around zero.

Figure 38

Overlayed density plots showing the distribution of ΔPeriodicEnergy_Max in contrast to ΔPeriodicEnergy_Min (top) and ΔPeriodicEnergy_Mean (bottom).



The contrast observed motivated further inspection of the speaker-level random effects on the models analysing ΔPeriodicEnergy_Min, ΔPeriodicEnergy_Mean and ΔPeriodicEnergy_Max. First, I extracted speaker-level random effects from all models with correlation test revealing that speaker-level random effects showed weak positive correlation between the results from the *minimum* and *maximum* values (Pearson's $r = 0.112$) as well as for the *mean* and *maximum* results (Pearson's $r = 0.254$).

The speaker-level effects (standard deviations) as extracted from the models' outputs confirm these weak correlations. Results from the models on minimum and mean values shows considerable speaker variation, especially for the use of editing terms ($SD = 0.09$ and $SD = 0.05$, respectively) and linguistic errors ($SD = 0.06$ and $SD = 0.04$, respectively). In contrast, the model on maximum periodic energy has minimal speaker-level variation (SD ≈ 0.00 to 0.01 across predictors) reinforcing that speaker differences are not robust on maximum periodic energy. which summarizes speaker-level random effects from all models, illustrates these differences and provides a brief interpretation on the variation. In addition, Figure 39 shows the contrasts on the density of the speaker-levels effects, illustrating the weak correlation between random effects from *mean* and *minimum* to those of the *maximum* values. Note that while speaker-level effects on ΔPeriodicEnergy_Min and ΔPeriodicEnergy_Mean are broadly spread, for ΔPeriodicEnergy_Max values are tightly clustered around zero. This suggests that while minimum and mean periodic energy more successfully captures phonetic contrasts between reparandum and repair, maximum periodic energy remains largely uninformative in distinguishing them.

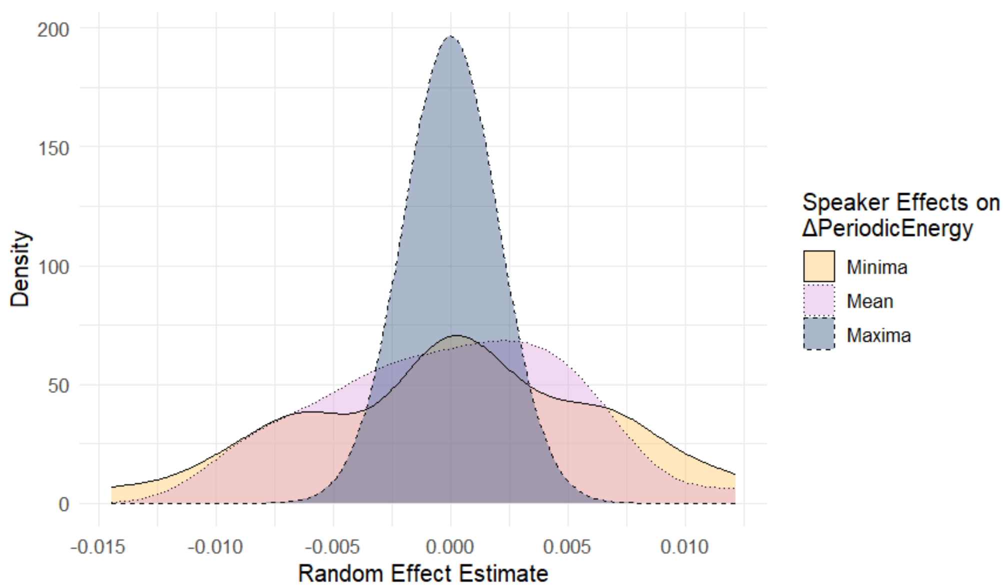
Table 24

Summary of speaker-level random effects for ΔPeriodicEnergy_Max, ΔPeriodicEnergy_Min and ΔPeriodicEnergy_Mean.

	Delta Periodic Energy	***SD***	***SE***	**Lower bound**	**Upper bound**	**Interpretation**
Intercept	*Minima*	0.03	0.02	0.00	0.07	*Speaker effects on Minima and Mean show variability (SD=0.03 and SD=0.02, respectively), while Maxima remains stable (SD=0.00).*
	Mean	0.02	0.01	0.00	0.05	
	Maxima	0.00	0.00	0.00	0.01	
Completeness Incomplete	*Minima*	0.04	0.03	0.00	0.12	*Completeness has notable speaker variation in both Minima and Mean (SD=0.04) but minor in Maxima (SD=0.01).*
	Mean	0.04	0.02	0.00	0.9	
	Maxima	0.01	0.01	0.00	0.03	
Editing Terms Yes	*Minima*	0.09	0.05	0.00	0.20	*Editing Terms show high speaker variation in Minima (SD=0.09)and moderate variability in Mean (SD=0.05) but very low in Maxima (SD=0.01).*
	Mean	0.05	0.03	0.00	0.12	
	Maxima	0.01	0.01	0.00	0.02	
Repair Type Factual	*Minima*	0.04	0.03	0.00	0.11	*Factual errors show moderate speaker variability in Minima (SD=0.04), less variability in Mean (S=0.02) and very low in Maxima (SD=0.01).*
	Mean	0.02	0.01	0.00	0.06	
	Maxima	0.01	0.00	0.00	0.01	
Repair Type Linguistic	*Minima*	0.06	0.05	0.00	0.18	*Linguistic errors show higher speaker variability in Minima (SD=0.06), moderate in Mean (0.04) but minor in Maxima (SD=0.01).*
	Mean	0.04	0.03	0.00	0.10	
	Maxima	0.01	0.01	0.00	0.02	

Figure 39

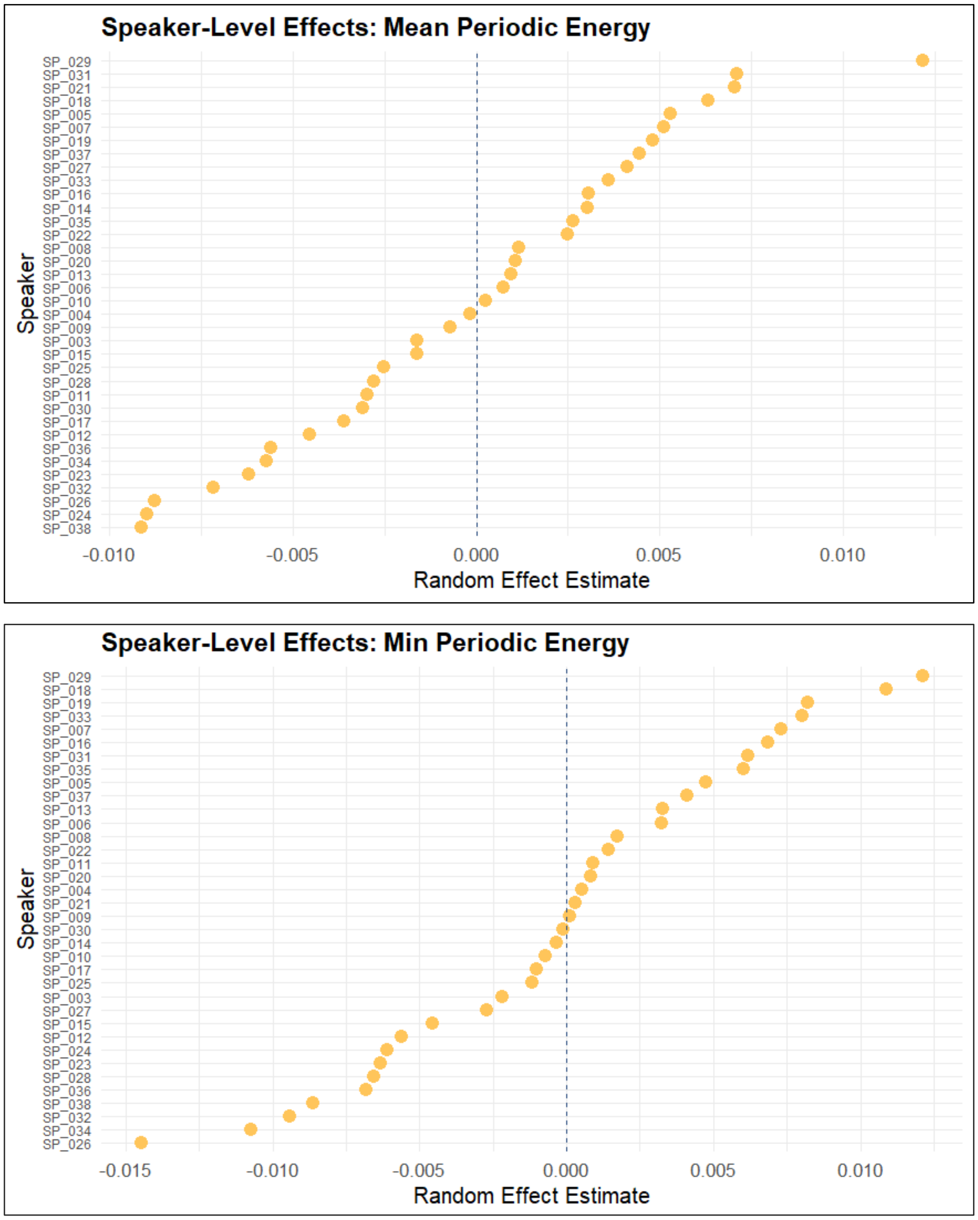
Overlay density plot comparing speaker-level random effects on ΔPeriodicEnergy_Max, ΔPeriodicEnergy_Min and ΔPeriodicEnergy_Mean.



As observed, speaker-related variation in the minimum and mean periodic energy values contrasts sharply with the relative stability of the maxima. The latter motivated further exploration into ΔPeriodicEnergy_Min and ΔPeriodicEnergy_Mean to observe how periodic energy changes by speaker within the two models fit for both measurements. Figure 40 illustrates the variation as discovered from the speaker-level effects, with some speakers systematically reducing minimum and mean periodic energy in their repairs (e.g., SP_026, SP_024 and SP_034) while others increase it (e.g., SP_018, SP_029 and SP_031), suggesting possible individualized prosodic marking strategies.

Figure 40

Speaker-Level Random Effects on ΔPeriodicEnergy_Mean (top) and ΔPeriodicEnergy_Min (bottom) per speaker.



It is important to attempt to describe how speaker-specific behaviours align, or diverge, from the broader trends identified in the predictor analysis reported earlier. For example, do all speakers consistently follow the general tendency for lower periodic energy in appropriateness repairs or in repairs with incomplete reparanda? Or do individual differences suggest alternative strategies? With that in mind, I will provide a closer examination of individual cases to contextualize these findings within prosodic behaviour as observed from periodic energy in actual speech.

To illustrate how speaker variation influences prosodic realization in repairs, I will present examples from speakers that showed estimate values of periodic energy in repairs at both ends of the distribution; that is, speakers who tend to strengthen their repairs versus speakers who tend to reduce energy (as seen in Figure 40). I will also include cases from speakers whose estimates where closer to zero, where no reduction or increase in energy was observed in their mean estimates. In the illustrations accompanying the selected examples (i.e., periograms), the shifts on periodic energy from reparandum to repair can be tracked through variations in the thickness and darkness of the blue curve, which also modulates F0. Furthermore, to aid visualizing periodic energy, we can inspect the red curve representing the mass, which is related to the prosodic strength of each periodic energy fluctuation (i.e., each syllable) inherently encoding periodic energy changes.

I will start with speakers who tend to reduce prosodic strength. Speaker 29 showed mean negative estimate values of periodic energy meaning their repairs tend to have lower strength in comparison to reparanda. Figure 41 and Figure 42 illustrate this behaviour by comparing reparanda and repairs in two different repair sequences. In the first example (Figure 41), the speaker changes their mind on judging an event. First, they describe the situation as ‘fair’ (*justo*, in Spanish) to latter change their mind and repair to ‘not totally fair’ (*no justo del todo)*, which constitutes a factual repair, produced after fully articulated (complete) reparanda. This repair exhibits a substantial decrease in mean periodic energy (-0.288) in comparison to the reparandum, while it is notable that maximum periodic energy remains nearly unchanged (0.001) in the two phases. In second example (Figure 42), the same speaker repairs from *salir* 'exiting' to *ser libre* 'being free', when referring to a person being released from prison. This case is an illustration of an appropriateness repair with incomplete reparanda, as only the first syllable of the reparandum is articulated. In the example, mean periodic energy is reduced (-0.398) and, again, the maximum energy reflects a negligible reduction (-0.041).

The same pattern of reduction is found in Speaker 38, whose estimated reduction is the highest among participants when observing mean values. In the example illustrated in Figure 43, the speaker repairs from one colour to another (*rosado* 'pink' to *gris* 'grey') while making a description. The move involved a reduction of -0.106 units in mean and also maximum periodic energy. Similarly, the speaker reduces the energy in the repair *rayitas* 'squiggles' that comes after the reparandum *escrito* 'text' (Figure 44). In this example, the speaker refers to an image that resembles a written text as such but after fully articulating the first choice, repairs to express that it is not actual text but a continuous line that curls and loops in an irregular way. There is energy reduction in the mean values of the repair (-0.105) as compared to the error, but the maximum values remained practically unchanged (0.004). The reduction is such, that even that final syllable of the repair, which has an open vowel in the nucleus, is weaken to a point where the whole syllable becomes unvoiced which results in no observed periodic energy[44].

---

[44] Although this reduction is easily observable in the repair's periograms, no measurement is extracted from that syllable since there is not periodic energy to be measured. However, values obtained from the first two syllables were enough to capture the reduction in the items' comparison.

To conclude the analysis of speakers who tend to reduce periodic energy in their repairs, Speaker 32 stands out as a constant reducer with 77.8% of their repairs showing negative values in both delta mean and minimum periodic energy. Nevertheless, I have chosen to illustrate a case of prosodic strengthening for this speaker, contrary to their predominant pattern, since our data also reveals instances of intra-speaker variation. Figure 45 presents an example of an appropriateness repair, with no editing terms, that is produced after a complete reparandum. The speaker is describing the location of an item in a picture and initially uses the diminutive *abajito* 'a little below'. This is then repaired to *abajo* 'below', producing the preposition in its standard form without additional modifications. The repair is marked by a notable increase in mean and minimum periodic energy (0.481 and 0.782, respectively), while the maximum is again nearly unchanged (0.005).

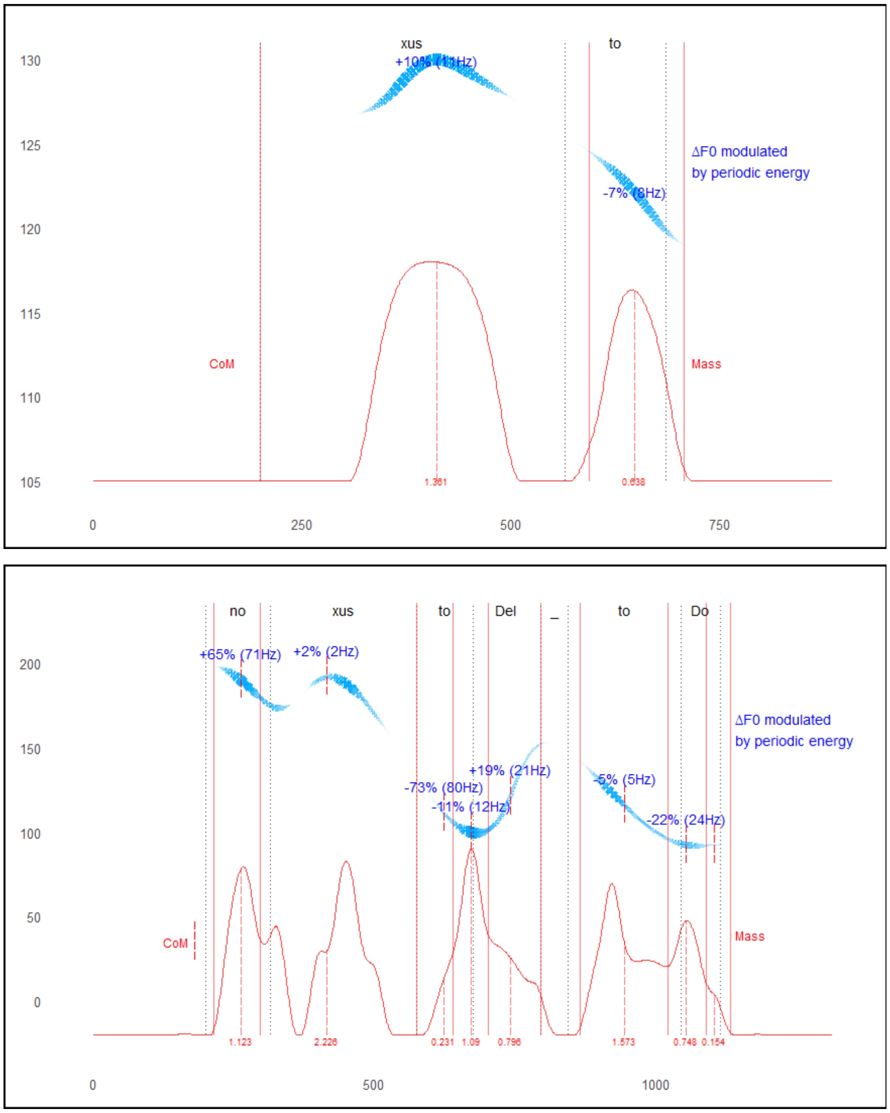
Moving the analysis to speakers that produce the opposite pattern, let us start by reviewing Speaker 29, who showed the highest values of increased periodic energy in our models. This pattern is evident in the repair *tres* 'three', which corrects *cuatro* 'four' (reparandum). A substantial increase in mean periodic energy is observed (0.340) in comparison to the reparandum, while maximum periodic energy remains nearly unchanged (-0.005) (see Figure 46). Similarly, Figure 47 illustrates a repair that corrects an issue of number agreement between the determiner and the noun, constituting a linguistic error repair in which no editing terms were present. In the reparandum, the speaker incorrectly uses the singular determiner *uno* 'one' to match the plural noun *disparos* 'shots', creating an ungrammatical structure since in Spanish determiners (such as quantifiers) must agree in gender and number with the noun they modify. In the repair, the plural determiner *unos* 'a few' is used instead, resolving the mismatch. This repair sequence led to an increase in periodic energy mean (0.295), which is also observable in the maximum values (0.329). Speaker 18 also tends to increase the energy in their repairs as observed in their estimate coefficients. One example of this pattern arises when the speaker confuses the names of two widely known broadcasting corporations, which led to a factual repair. The speaker also quickly acknowledges the error by articulating the editing term '*no*' in between the error and the correction. The repair 'Caracol', pronounced as [ˌka.ɾa.ˈkol] shows a moderate increase in mean periodic energy (0.173) when compared to the reparandum 'RCN', pronounced as [e.ˌre.se.ˈe.ne]. Figure 48 illustrates the pattern.

To finish the illustration on the contrast in periodic energy in repair sequences, I will present examples extracted from two speakers who are more neutral in their estimates but from whom we can also observe some reduction in energy in the repair in two different types of repairs. The first one is an appropriateness repair that is produced after complete reparanda by Speaker 27. In the sequence the speaker shifts from the noun *bolas* 'balls' (reparandum) to *balones* 'sport balls' (repair), likely to avoid the double meaning of balls. The term *balones* is more specific and typically used in sports contexts, reducing any unintended interpretations. This repair sequence, illustrated in Figure 49 did not include editing terms and had a complete reparandum. It involved a reduction of periodic energy in the repair mean values (-0.154) and some reduction was also observable in the maximum energy (-0.083). The other examples correspond to Speaker 33, who also produced a repair sequence in which a reduction is observed in the repair phase. In this example on top on editing terms, the speaker makes a clear reference to the trouble source, addressing the issue directly and even laughing at it after repairing. This behaviour is what we earlier refer to as ATTENTION TO ERROR. Repairs with this structure were not included in the analysis of temporal organization because of the effects these comments had on timing the repair, however, these cases were not excluded from the present analysis because we could account for the prosody of these repairs irrespectively of the fact that the trouble source was addressed directly since that did not affect the observed measurements. Having explained that, the sequence aims at repairing the noun *pollito* ‘little chick’ (reparandum), which is only partially articulated, to the more general noun *pajarito* ‘little bird’. This factual repair showed a reduction in periodic energy mean (-0.299) while the maximum showed similar strength in both phases, reducing only a few units (-0.034). In a similar way to other illustrations, the strength reduction is such that voicing almost disappears in the unstressed syllables of the repair *pajarito*. This becomes visible in the corresponding periograms, which are available in Figure 50, illustrating the contrast.

Overall, the findings suggest that *mean* and *minimum* capture different prosodic aspects of speech repairs than *maximum* periodic energy. While mean periodic energy varies significantly across speakers, maximum periodic energy remains more stable, indicating that these two measures may reflect distinct phonetic mechanisms. To conclude this section and provide a visual summary of the reported periodic energy changes. The illustrations below present the periograms for all analysed repair sequences, organized by speaker and repair type. These figures demonstrate the observed reduction and increase in periodic energy, reinforcing the patterns discussed above.

Figure 41

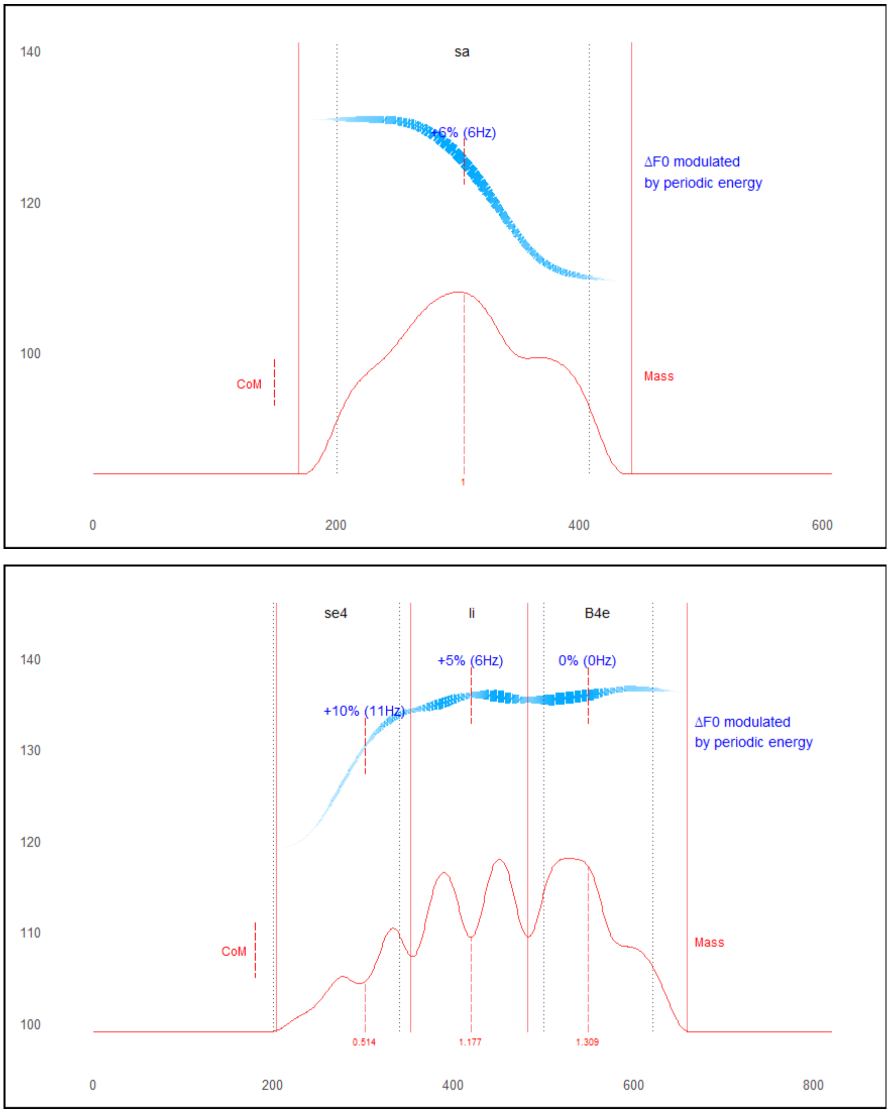
Periograms of the Reparandum 'justo' (top) and Repair 'justo del todo' (bottom) as produced by Speaker 26.



Note. The periograms illustrate the periodic energy dynamics of this factual repair. A reduction in periodic energy is observed in the repair, as indicated by the lighter and thinner blue curve in comparison to the reparandum. This supports the trend that factual repairs can also exhibit reduced periodic energy, though variation exists across speakers.

Figure 42

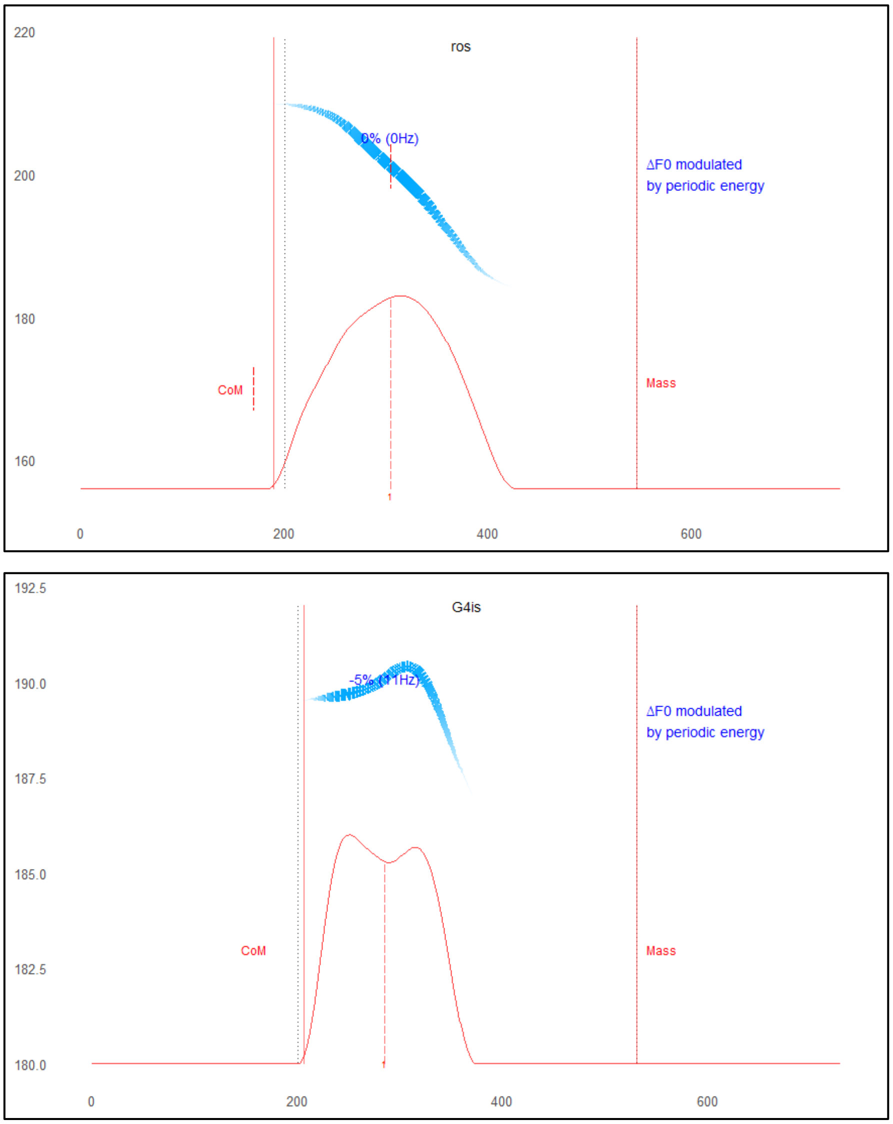
Periograms of the incomplete Reparandum 'salir' (top) and Repair 'ser libre' (bottom) as produced by Speaker 26.



Note. The periograms illustrate the periodic energy dynamics of an appropriateness repair. A substantial reduction in periodic energy is observed in the repair, as indicated by the lighter and thinner blue curve in comparison to the reparandum. This pattern aligns both the tendency for incomplete-reparanda repairs to have reduced energy and also with the broader trend of appropriateness repairs exhibiting reduced periodic energy.

Figure 43

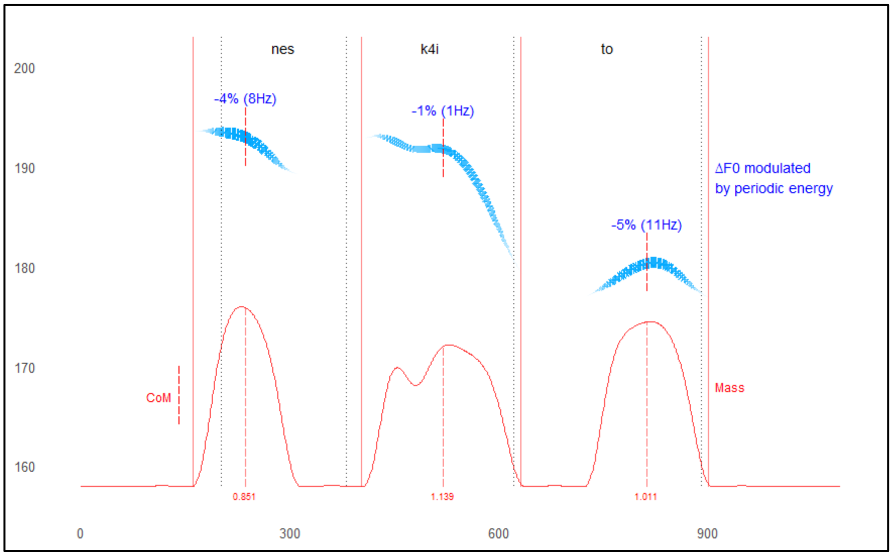
Periograms of the incomplete Reparandum 'rosada' (top) and Repair 'gris' (bottom) as produced by Speaker 38.

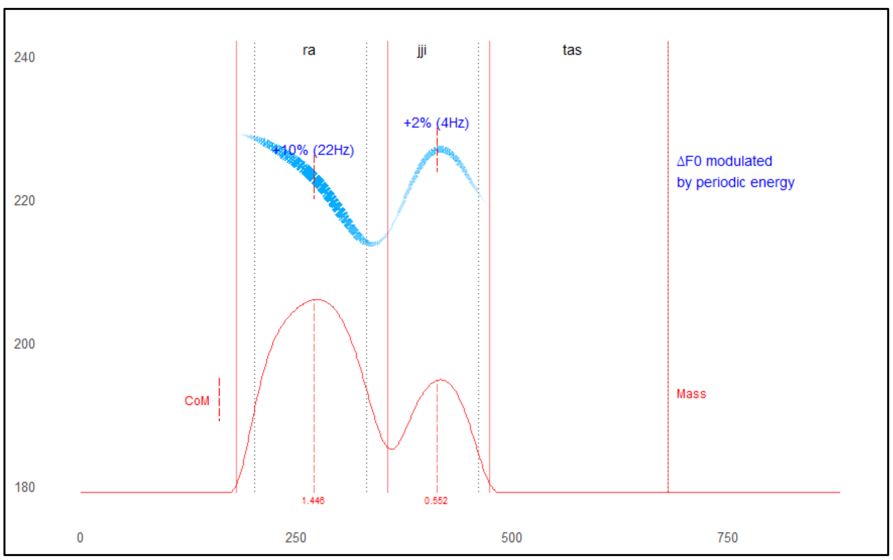


Note. The periograms illustrate the periodic energy changes in a factual, incomplete repair with an editing term. The repair demonstrates a reduction in periodic energy, as seen in the lighter and thinner blue curve compared to the reparandum. This pattern is consistent with the observed trend, where factual repairs, resulting from incomplete reparanda tend to exhibit lower periodic energy, reinforcing their prosodic distinctiveness.

Figure 44

Periograms of the Reparandum 'escrito' (top) and Repair 'rayitas' (bottom) as produced by Speaker 38.

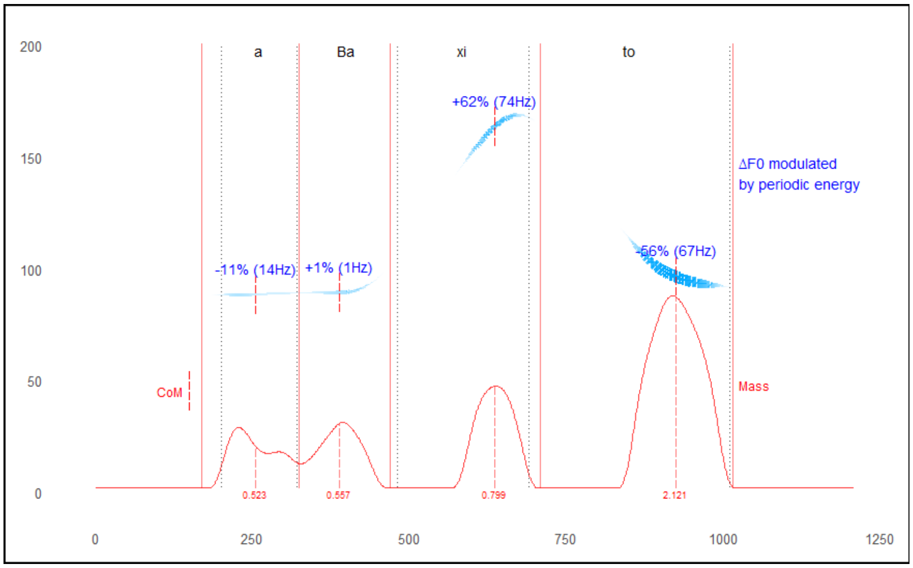


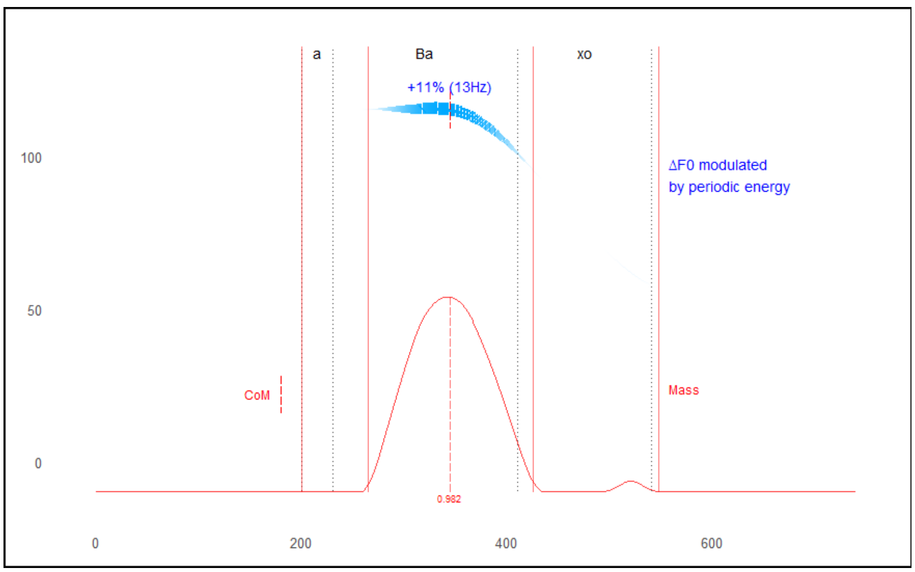


Note. The periograms illustrate the periodic energy changes in a factual, incomplete repair with an editing term. The repair demonstrates a reduction in periodic energy, as seen in the lighter and thinner blue curve compared to the reparandum. This pattern is consistent with the observed pattern, where factual repairs, resulting from incomplete reparanda tend to exhibit lower periodic energy, reinforcing their prosodic distinctiveness.

Figure 45

Periograms of the Reparandum 'abajito' (top) and Repair 'abajo' (bottom) as produced by Speaker 32.

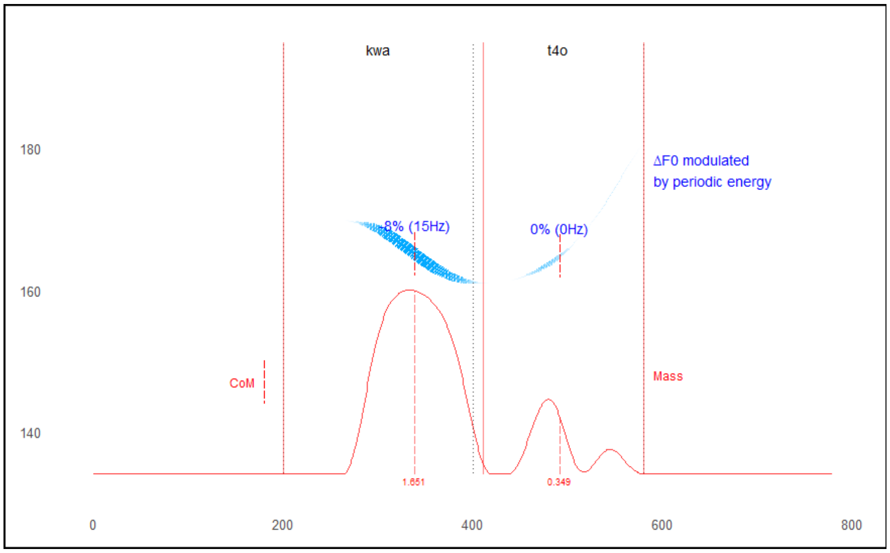


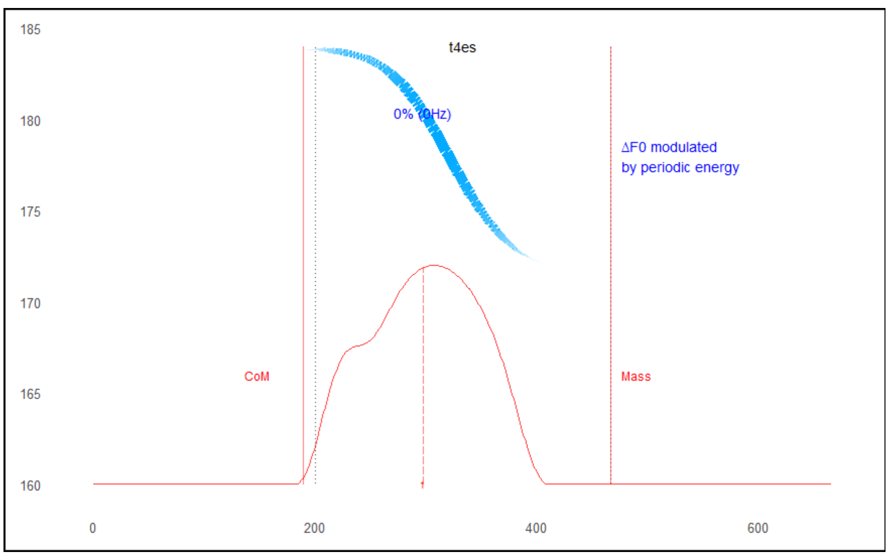


Note. The periograms illustrate the periodic energy changes in an appropriateness repair. The repair exhibits a reduction in periodic energy, evidenced by the lighter and thinner blue curve compared to the reparandum. This pattern aligns with the general tendency of appropriateness repairs to have lower periodic energy, reinforcing their prosodic characteristics.

Figure 46

Periograms of the Reparandum 'cuatro' (top) and Repair 'tres' (bottom) as produced by Speaker 29.

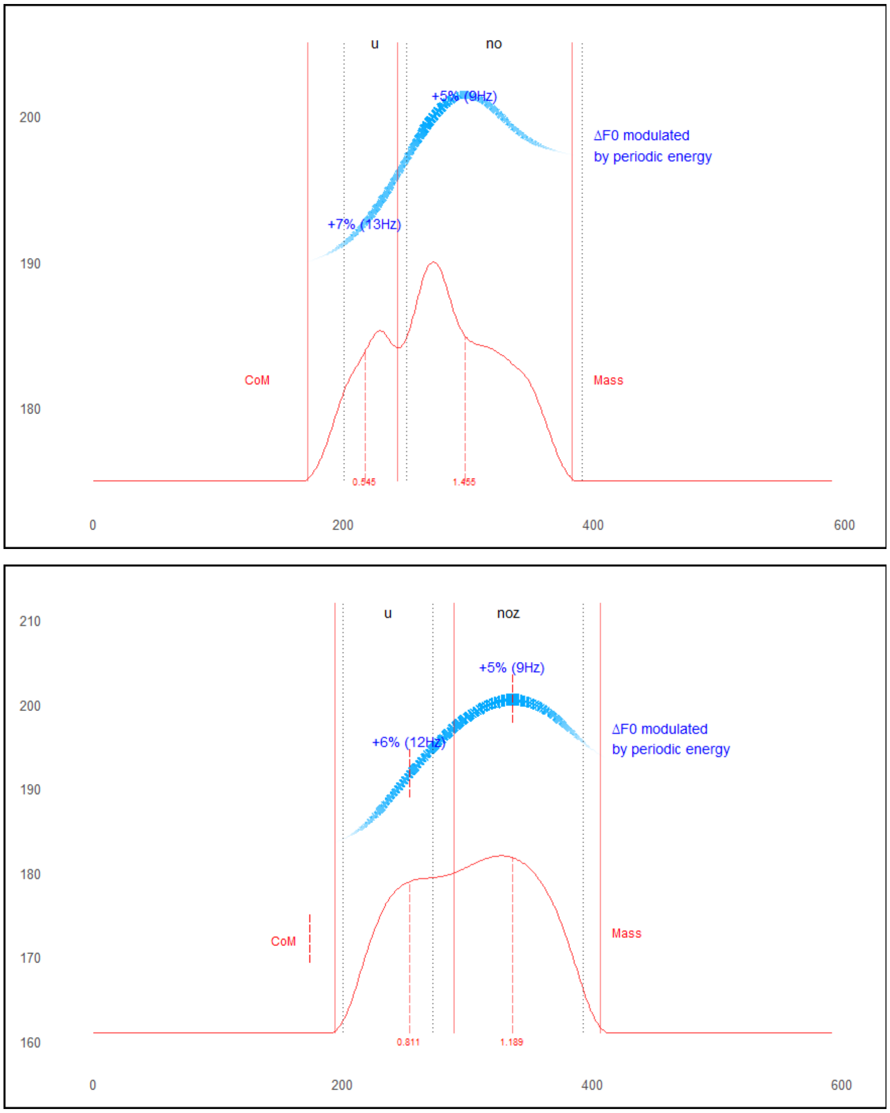




Note. The periograms illustrate the periodic energy changes in a factual, complete repair with no editing terms. The repair demonstrates how we can also have increased mean periodic energy in repair since the repair phase exhibits a thicker and darker blue F0 curve compared to the reparandum.

Figure 47

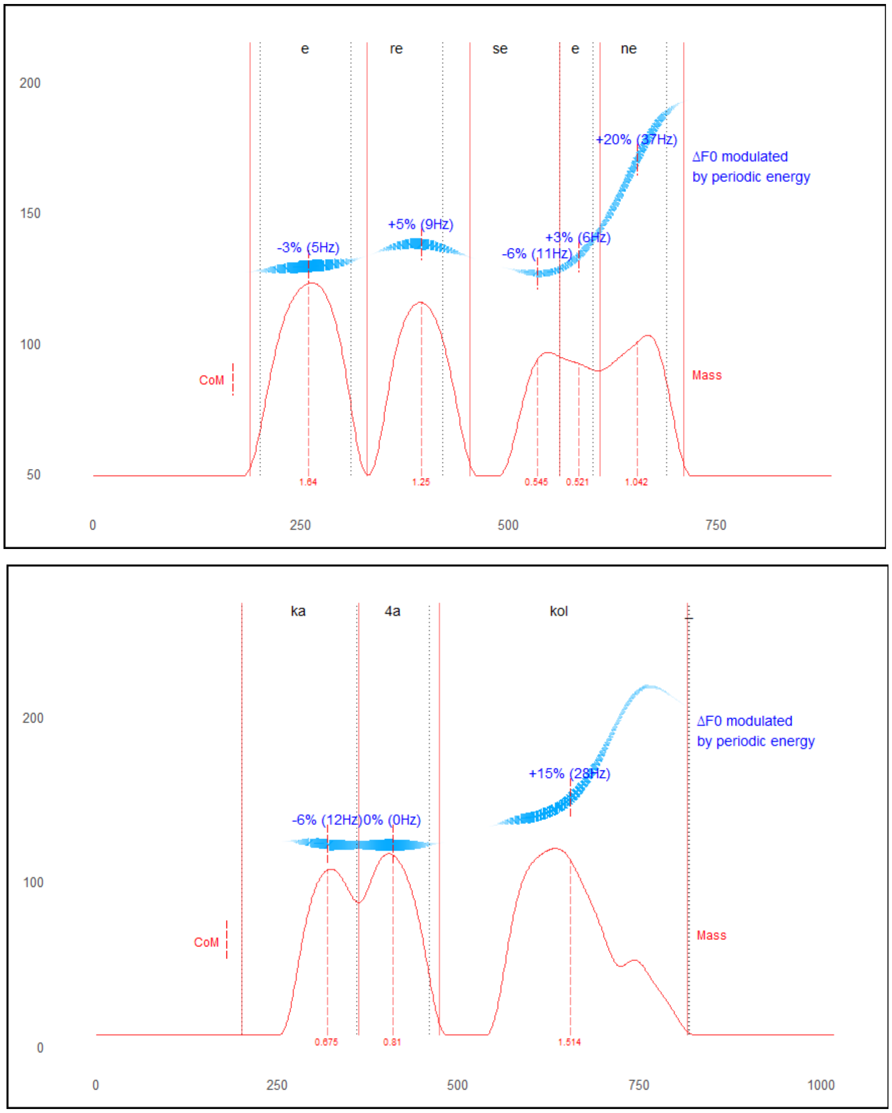
Periograms of the Reparandum 'uno' (top) and Repair 'unos' (bottom) as produced by Speaker 29.



Note. The periograms illustrate a moderate increase in periodic energy in this linguistic repair. The repair phase exhibits a thicker and darker blue F0 curve compared to the reparandum, signalling an increase in periodic energy. This case is particularly relevant as it represents the only linguistic repair in the dataset, highlighting that linguistic errors may also be prosodically reinforced depending on speaker strategies.

Figure 48

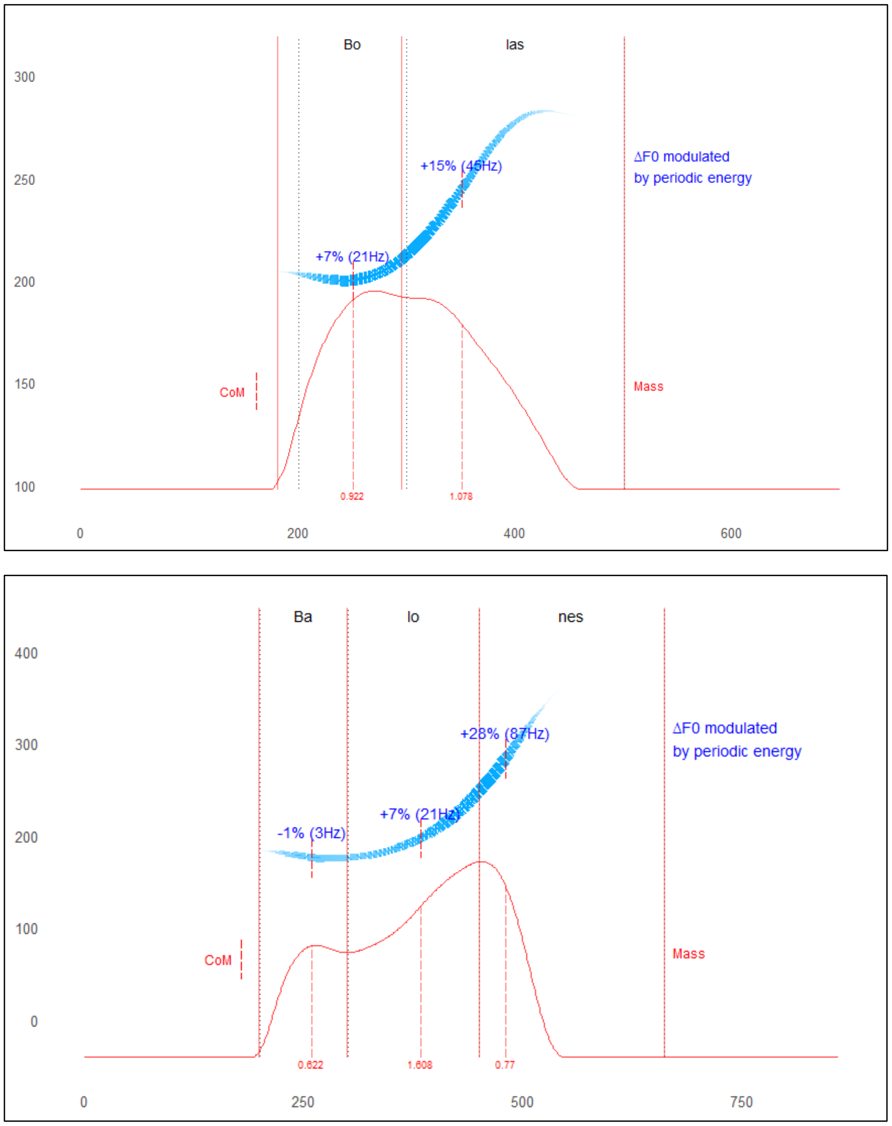
Periograms of the Reparandum 'RCN' (top) and Repair 'Caracol' (bottom) as produced by Speaker 18.



Note. The periograms illustrate an increase in periodic energy in this factual repair. The repair phase exhibits a thicker and darker blue F0 curve compared to the reparandum, signalling a strengthening of periodic energy. This case contrasts with the general pattern observed in some factual repairs that reduce periodic energy, highlighting speaker-specific prosodic strategies in repair realization.

Figure 49

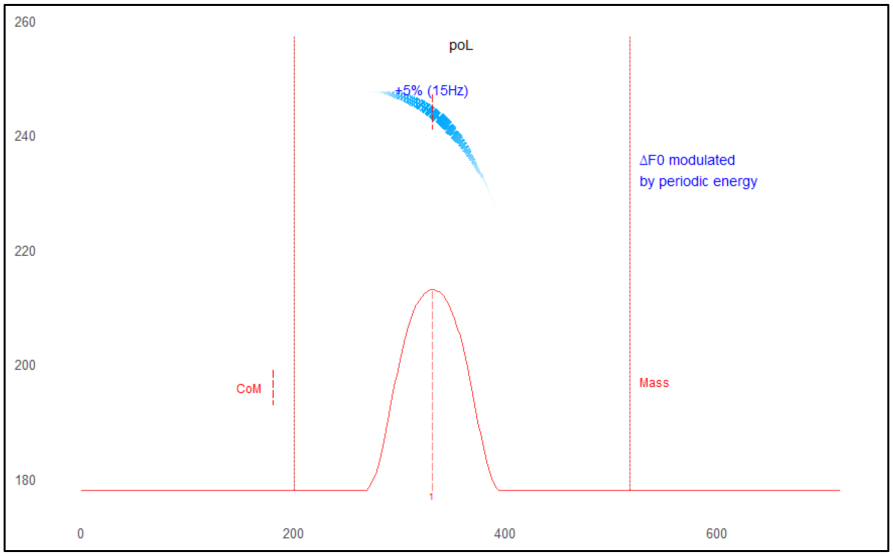
Periograms of the Reparandum 'Bolas' (top) and Repair 'Balones' (bottom) as produced by Speaker 27.

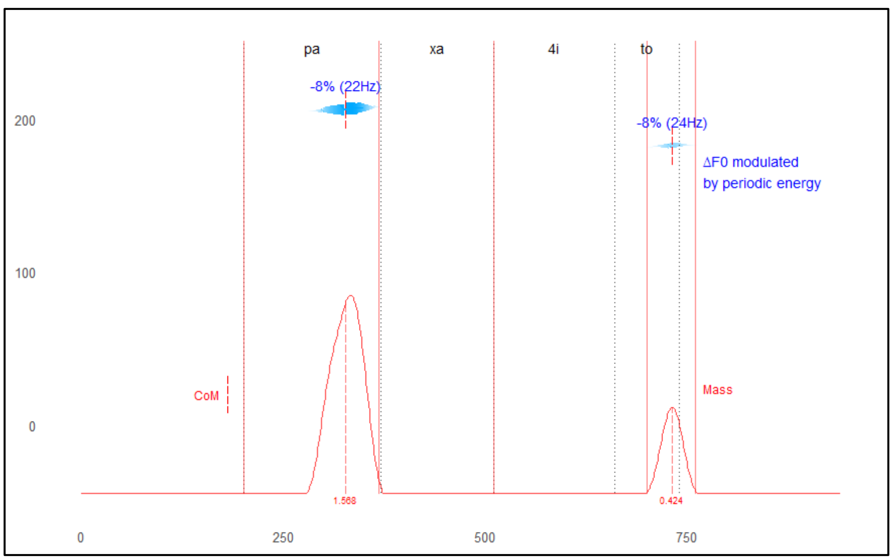


Note. The periograms illustrate a reduction in periodic energy in this appropriateness repair. The repair segment is marked by a thinner and lighter F0 contour relative to the reparandum, indicating decreased voicing strength. This case aligns with the general pattern observed in appropriateness repairs, where speakers tend to reduce prosodic prominence, possibly as a strategy to downplay the repair within the conversational flow.

Figure 50

Periograms of the Reparandum 'pollito' (top) and Repair 'pajarito' (bottom) as produced by Speaker 18.





Note. The periograms illustrate an increase in periodic energy in this factual repair. The repair phase exhibits a thicker and darker blue F0 curve compared to the reparandum, signalling a strengthening of periodic energy. This case contrasts with the general pattern observed in some factual repairs that reduce periodic energy, highlighting speaker-specific prosodic strategies in repair realization.

4.4.2 Relevance of Completeness, Editing Terms and Repair Type on Pitch

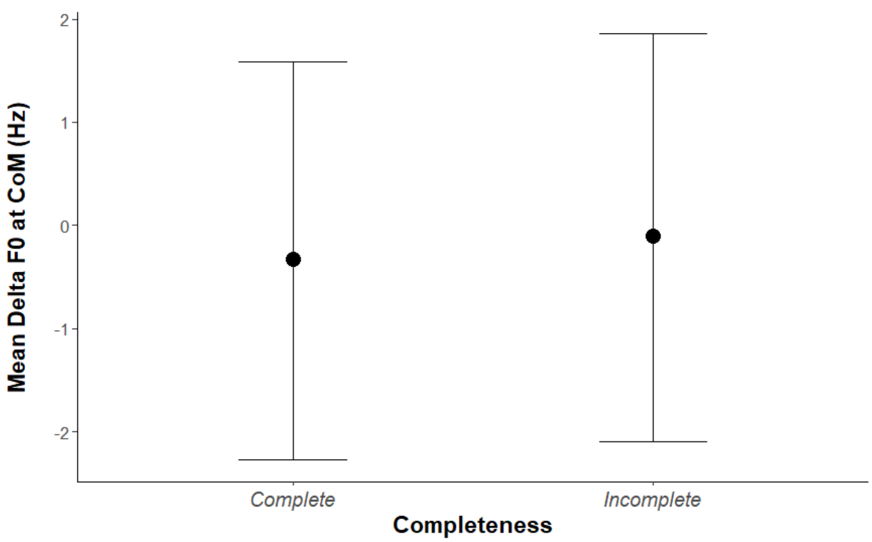
As reported earlier, ΔF0atCoM_Mean showed a strong positive relationship with both ΔF0atCoM_Max (r = 0.839, $p < 0.001$) and ΔF0atCoM_Min (r = 0.821, $p < 0.001$). Additionally, both the absolute and relative measurements of Dynamic F0 were highly correlated (r = 0.934, $p < 0.001$). Consequently, in this section I will report on results as extracted from the model analysing ΔF0atCoM_Mean, directly accounting pitch shifts in Hertz (Hz). Secondly, I will report from the model analysing ΔDynamicF0_Abs, where the findings will also be presented in Hertz, which keeps consistency with the report on F0 at CoM.

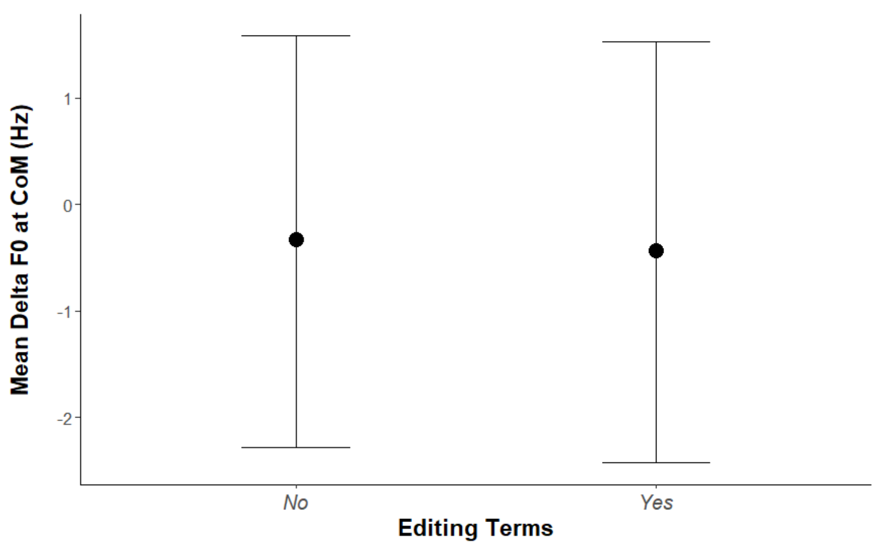
Before moving into the results, an additional correlation test was conducted between ΔF0atCoM_Mean and ΔDynamicF0_Abs, being these were the selected measurements to account for pitch variation in the data set. The results point to a moderate-to-strong positive correlation between the two (r = 0.648, $p < 0.001$), suggesting that as the mean F0 at CoM increases, the F0 absolute change also tends to increase. The 95% confidence interval [0.589, 0.702] confirms that the observed relationship on pitch shifts is unlikely due to a random chance.

Let us start by reporting on the results of our predictors on the mean F0 at CoM. The intercept represents appropriateness repairs when the reparandum is complete and contains no editing terms. The estimate (-0.33, *SE* = 0.98, 95% CI [-2.28, 1.59]) suggests a decrease in pitch, but the wide credible interval indicates high uncertainty. This means that while a pitch reduction is possible, it is not statistically robust, and the true effect could be small, large, or even non-existent. On *Completeness*, results show a small increase in pitch on the repairs compared to reparanda when the reparandum was incomplete, with an estimate of 0.23 (*SE* = 0.48, 95% CI [-0.70, 1.17]). Moving on to *Editing Terms*, the analysis reveals no major difference in pitch between reparandum and repair based on the presence or absence of editing expressions. Repair sequences that included editing terms showed a small pitch reduction, approximately -0.11 (*SE* 0.48 and a CI [-1.05, 0.85]) compared to sequences with no editing expressions. Lastly, regarding *Repair Type*, appropriateness repairs exhibit a small decrease in pitch but with high uncertainty. Factual errors also exhibited a small decrease in pitch, -0.07 (*SE* = 0.48 and a 95% CI: [-1.05, 0.85]). In contrast, linguistic errors showed a small pitch increase, suggesting than they tend to have a higher pitch than appropriateness and factual repairs, with an estimate of 0.07 (*SE* = 0.49, and a 95% CI [-0.89, 1.05]. As can be seen, credible intervals for all three predictors include zero, suggesting that none of these effects are statistically robust. That indicates that the observed shifts in repairs pitch levels based on *Completeness*, the presence or absence of *Editing Terms* and the semantic *Repair Type* are negligible. Figure 51 illustrates the impact of the different predictors on ΔF0atCoM_Mean. Also, a full summary of estimates for the variables *Completeness*, *Editing Terms* and *Repair Type* on ΔF0atCoM_Mean is given in Table 25 which includes a brief interpretation of the findings.

Figure 51

Conditional effects plot for Completeness (top), Editing Terms (middle) and Repair Type (bottom) as extracted from the student-t model analysing ΔF0atCoM_Mean. The error bars display 95% credible intervals; the dots represent posterior means.





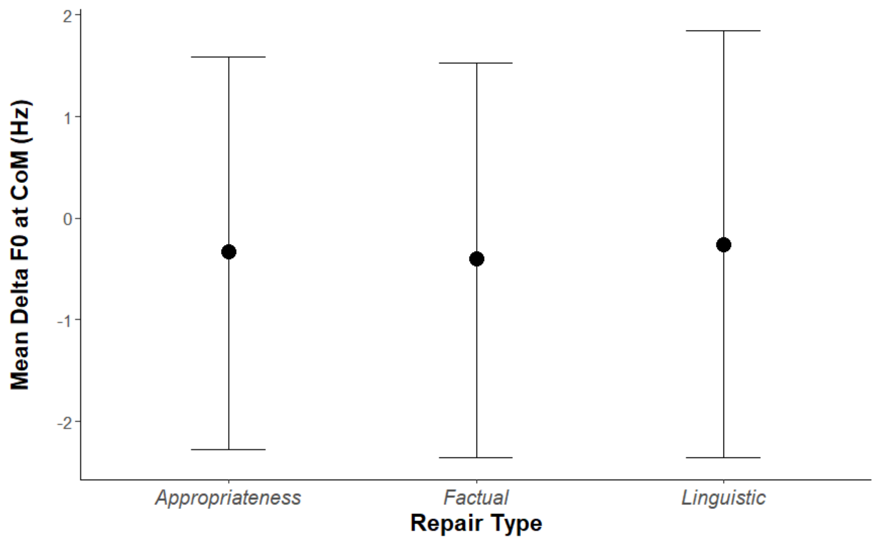


Table 25

Posterior mean, standard error, 95% credible interval and brief interpretation for each predictor in the model analysing ΔF0atCoM_Mean

	Estimate	*SE*	Lower bound	Upper bound	Interpretation
Intercept	-0.33	0.98	-2.28	1.59	*Baseline periodic energy for complete repairs, appropriateness repairs, without editing terms.*
Completeness Incomplete	0.23	0.48	-0.70	1.17	*Repairs with* incomplete *reparanda have a higher pitch by 0.23 Hz compared to complete ones, but the effect is not robust.*
Editing Terms Yes	-0.11	0.48	-1.05	0.85	*The is a small reduction (-0.11 Hz) in pitch in repairs with* editing terms*, but with uncertainty.*
Repair Type Factual	-0.07	0.48	-1.01	0.87	*Factual errors also showed reduction in pitch by 0,07 Hz, indicating lower pitch than in appropriateness repairs but the effect is not robust.*
Repair Type Linguistic	0.07	0.49	-0.89	1.05	*Linguistic errors increase pitch by 0.07 Hz, making them stronger than appropriateness and factual repairs but with uncertainty.*

Before moving into the next set of results, it is relevant to acknowledge that to maintain consistency with the other analyses performed and reported, for ΔF0atCoM_Mean, I also explored for interactions between predictors. However, no strong evidence of a robust relationship between *Completeness*, *Editing Terms* or Repair Type was found; therefore, no further interaction effects will be reported.

Moving on to results on Dynamic F0, similarly to the report on F0 at CoM, results on ΔDynamicF0_Abs suggest no robust effects for any of the independent variables. Credible intervals for all three predictors cross zero, indicating lack of strong evidence for a reliable effect of either *Completeness*, *Editing Terms* or *Repair Type*. Nevertheless, I will summarize the observed trends below.

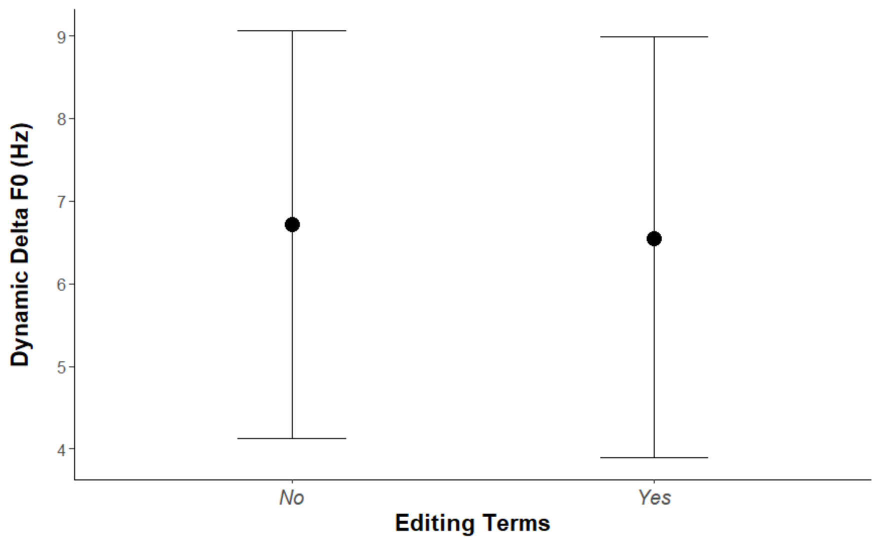
The intercept represents appropriateness repairs when the reparandum is complete and contains no editing terms. The estimate (6.69, *SE* = 1.25, 95% CI [4.13, 9.06]) suggests that, on average, appropriateness repairs exhibit a higher upward shift in pitch. However, given the wide confidence interval, there is still some uncertainty regarding the exact magnitude of this effect. For *Completeness*, results suggest that repairs with incomplete reparanda tend to have a greater upward pitch shift in the repair item. Results show an estimate upward shift approximately 0.43 Hz higher in repairs with incomplete reparanda (*SE* = 0.49, 95% CI [-0.54, 1.38]) in contrast to repair sequences which reparanda was fully articulated. For *Editing Terms*, results indicate a slightly smaller pitch shift in repair items compared to their reparanda, when editing terms are present, with a mean estimate of -0.17 Hz (*SE* = 0.49, 95% CI [-1.13, 0.78]). On *Repair Type*, factual errors showed a small decreased with a coefficient (-0.25 Hz; *SE* = 0.49 Hz, 95% CI [-1.21, 0.71]) while linguistic errors exhibited a slight increase (0.02 Hz; *SE* = 0.49 Hz, 95% CI [-

0.94, 0.99]). This means that while for linguistic errors and appropriateness infelicities the F0 upward shift tends to be higher in repairs, for factual errors the shift is smaller.

It is important to keep in mind that since the credible intervals for all predictors include zero, the evidence for a reliable effect remains inconclusive. Lastly on Dynamic F0, despite exploring interactions, no strong evidence of a robust relationship between predictors was found; therefore, no further interactions are reported. Figure 52 below, illustrates the impact of the different semantic repair subtypes on ΔDynamicF0_Abs. Additionally, full summary of estimates for the variables *Completeness*, *Editing Terms* and *Repair Type* on is given in Table 26.

Figure 52

Conditional effects plot for Completeness (top), Editing Terms (middle) and Repair Type (bottom) as extracted from the student-t model analysing ΔDynamicF0_Abs. The error bars display 95% credible intervals; the dots represent posterior means.


9
8
7
6
5
4
Dynamic Delta F0 (Hz)
No
Yes
Editing Terms

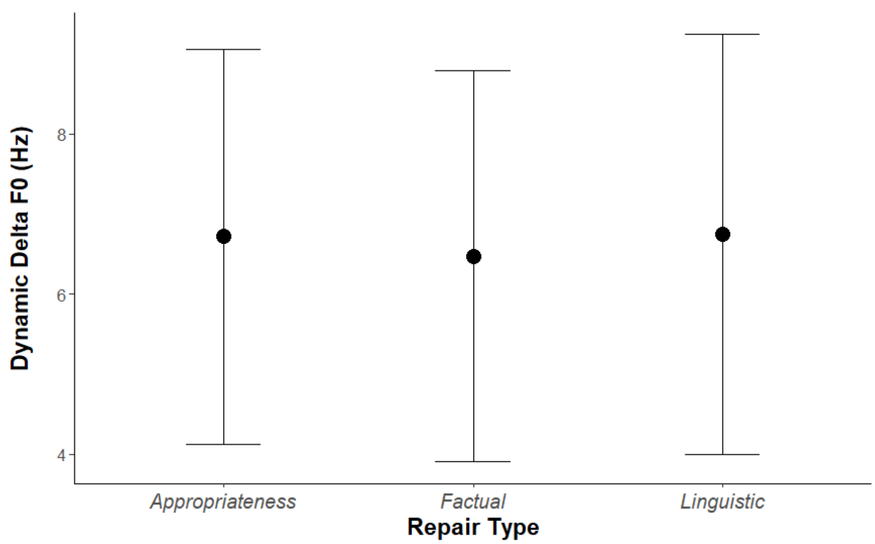

8
6
4
Dynamic Delta F0 (Hz)
Appropriateness
Factual
Linguistic
Repair Type

Table 26

Posterior mean, standard error, 95% credible interval and brief interpretation for each predictor in the model analysing ΔDynamicF0_Abs.

	Estimate	***SE***	**Lower bound**	**Upper bound**	**Interpretation**
Intercept	6.69	1.25	4.13	9.06	*Baseline periodic energy for complete repairs, appropriateness repairs, without editing terms.*
Completeness Incomplete	0.43	0.49	-0.54	1.38	*Repairs with* incomplete *reparanda have a higher pitch by 0.43 Hz compared to complete ones, but the effect is not robust.*
Editing Terms Yes	-0.17	0.49	-1.13	0.78	*The is a small reduction (-0.17 Hz) in pitch in repairs with* editing terms*, but with uncertainty.*
Repair Type Factual	-0.25	0.49	-1.21	0.71	*Factual errors showed reduction in pitch by 0,25 Hz, indicating lower pitch than in appropriateness repairs but the effect is not robust.*
Repair Type Linguistic	0.02	0.49	-0.94	0.99	*Linguistic errors increase pitch by 0.02 Hz, indicating higher pitch than appropriateness and factual repairs but with uncertainty.*

The findings from both models (i.e., ΔF0atCoM_Mean and ΔDynamicF0_Abs) show high consistency across predictors, reinforcing the positive correlation between these two prosodic measurements. In both cases, credible intervals include zero, indicating that the observed patterns lack strong statistical support. However, the general trends remain aligned. For instance, in both models, *Completeness* showed a weak tendency for incomplete repairs to exhibit slightly higher pitch values than complete ones. Similarly, the presence of editing terms was associated with a minor reduction in pitch, but again, this effect was small and uncertain. Regarding *Repair Type*, appropriateness and linguistic repairs tended to exhibit larger pitch increases, while factual repairs displayed smaller shifts, a pattern that held on both F0 models.

Summing up, the strong correlation between these two measurements and the alignment of the general trends observed suggest that they capture similar underlying prosodic patterns when analysing pitch changes.

4.4.3 Relevance of Completeness, Editing Terms and Repair Type on Speech Rate

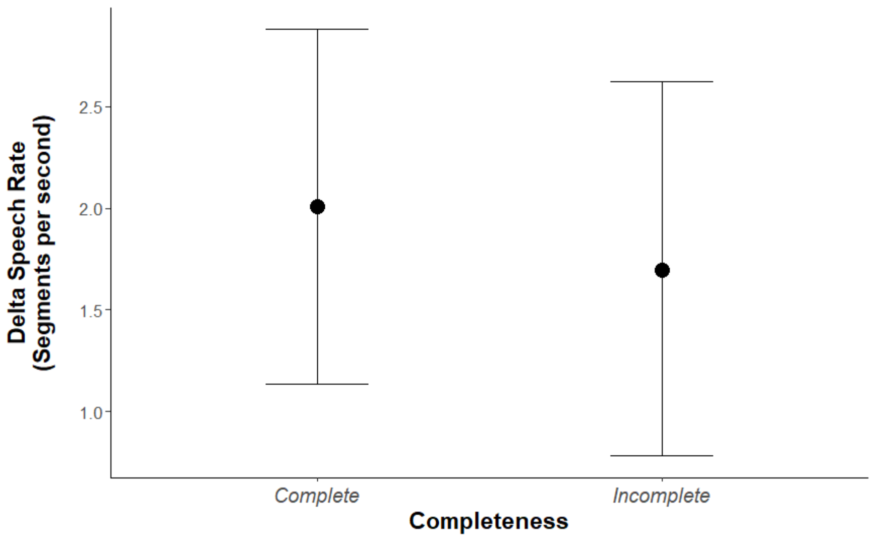
In this section I will report on results as extracted from the model analysing the contrast in articulatory speech rate in the repair and the reparandum. The analysis is based on counts of segments per second (i.e., ΔSpeechRate_Seg) as the primary measure of speech rate. As noted earlier, both speech rate measurements (segments per second and syllables per second) were highly positively correlated ($r = 0.823$, $p < 0.001$). Given this strong relationship, I chose to report on segments per second as lexical items in a repair sequence can vary in their phonotactic structure, resulting in different syllable types in terms of constituents and numbers. In addition, reparanda can be incomplete, meaning that some repairs may consist of a single syllable or even just a few segments, making segment-based comparisons more directly interpretable than syllable-based ones. Still, given the high correlation between the two measurements, it is expected that as the number of segments increases, the number of syllables does too. This report corresponds to the last acoustic measurement analysed to account for prosodic marking in repair sequences within the data set.

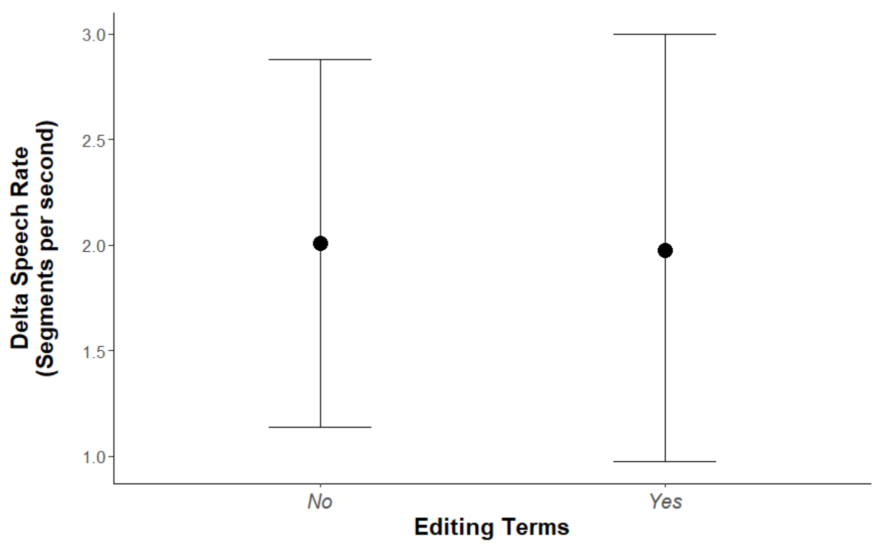
In the results, the intercept represents appropriateness repairs when the reparandum is complete and contains no editing terms. The estimate 2.01 (-0.33, *SE* = 0.45, 95% CI [1.13, 2.88]) suggests an increase of approximately two segments in repairs with the given configuration. Regarding *Completeness*, results show a decrease in speech rate in repairs compared to reparanda when the reparandum is incomplete, with an estimate of -0.31 (*SE* = 0.39, 95% CI [-1.08, 0.47]). However, since the credible interval includes a zero, this suggests that the effect is not robust, indicating that the observed reduction in speech rate for repairs with incomplete reparanda is negligible. Moving on to *Editing Terms*, the analysis reveals no major difference in speech rate between the reparandum, and repair based on the presence or absence of editing expressions. Repair sequences including editing terms exhibited a small decrease in speech rate of approximately -0.03 (*SE* = 0.41 and a CI [-0.83, 0.78]) compared to sequences without editing expressions. Again, the credible interval crosses zero, suggesting that the effect is statistically uncertain, meaning it could be negative, close to zero or even slightly positive. Finally, regarding *Repair Type*, both factual and linguistic errors exhibited a decrease in speech rate in the repair compared to the reparandum. Factual repairs showed a reduction of approximately -0.12 segments per second (*SE* = 0.40 and a 95% CI: [-0.90, 0.65]) while linguistic errors showed an even smaller reduction that is close to -0.02 segments per second (*SE* = 0.44 and a 95% CI: [-0.89, 0.85]).

Since the credible intervals for all three predictors reported above include zero, none of these effects can be considered statistically robust. That indicates that the observed differences in the speech rate are negligible, with *Completeness*, the presence or absence of editing terms and *Repair Type* showing no strong influence on the reported speech rates changes. Figure 53 illustrates the impact of the different predictors on ΔSpeechRate_Seg. Also, a full summary of estimates for the variables *Completeness*, *Editing Terms* and *Repair Type* on ΔSpeechRate_Seg is given in Table 26.

Figure 53

Conditional effects plot for Completeness (top), Editing Terms (middle) and Repair Type (bottom) as extracted from the student-t model analysing ΔSpeechRate_Seg. The error bars display 95% credible intervals; the dots represent posterior means.





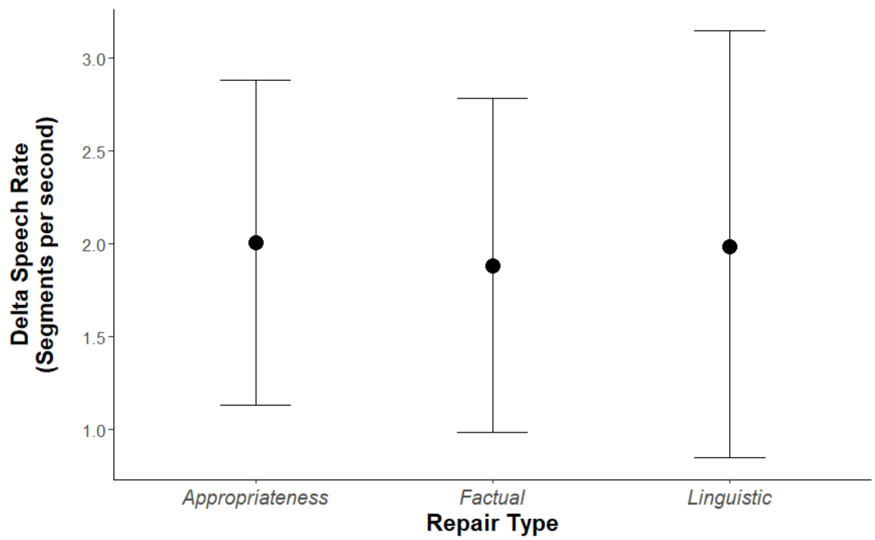


Table 27

Posterior mean, standard error, 95% credible interval and brief interpretation for each predictor in the model analysing ΔSpeechRate_Seg.

	Estimate	***SE***	**Lower bound**	**Upper bound**	**Interpretation**
Intercept	2.01	0.45	1.13	2.88	*Baseline periodic energy for complete repairs, appropriateness repairs, without editing terms.*
Completeness Incomplete	-0.31	0.39	-1.08	0.47	*Repairs with* incomplete *reparanda have lower rate by -0.31 segments per second compared to complete ones, but the effect is not robust.*
Editing Terms Yes	-0.03	0.41	-0.83	0.78	*There is a small reduction (-0.03 segments per second) in speech rate in repairs with* editing terms*, but with uncertainty.*
Repair Type Factual	-0.12	0.40	-0.90	0.65	*Factual errors also showed reduction (0,12 segments per second), indicating lower speech rate than in appropriateness repairs but the effect is not robust.*
Repair Type Linguistic	-0.02	0.44	-0.89	0.85	*Linguistic errors also showed reduction (0,02 segments per second) indicating lower speech rate than in appropriateness repairs but with uncertainty.*

Finally, after exploring significant interactions between the predictors, no robust evidence of a robust relationship between *Completeness, Editing Terms* or *Repair Type* was found. Consequently, no further interaction effects are reported.

4.5 Discussion

This chapter investigated prosodic marking in lexical self-repair by analysing periodic energy, pitch and speech rate in speakers' productions of both reparandum and repair within repair sequences that included different repair configurations. From the analysis reported in this chapter, it emerged that in Colombian Spanish spontaneous speech, speakers tend to lower periodic energy in their repairs, with repairs also showing higher pitch and a faster articulation speed. Chapter 4 also highlighted that neither *Completeness*, *Editing Terms* nor the semantic *Repair Type* predicted the speech rate or pitch levels for repair; however, analyses on periodic energy did show that repairs with incomplete reparanda were produced with lower strength than their fully articulated counterparts.

In relation to HYPOTHESIS A, which predicts hypoarticulation in repair, it was confirmed that repairs are more frequently spoken at a faster tempo, or *hypospeech*, with a mean estimate for speech rate that is 1.57 segments per second higher in repairs than in reparanda. This finding directly supports Plug's previous reports on fast tempo in self-initiated, self-repair in free conversations, as it emerged in spontaneous Dutch (Plug, 2011). Additionally, pitch was also found to be higher in repairs than in reparanda (9.60 Hz higher, mean). These findings are partially consistent with Plug's definition of marking being a combination between getting the job of correcting done fast, by hypo-articulating, while also emphasizing the corrected talk by increasing pitch and intensity (Plug, 2014, p. 13). We found both hypospeech and increased pitch in repairs but not increased loudness in our analysis of periodic energy. We cannot discard that an increase in both pitch and intensity could possibly be the case for our data as well; however, regarding periodic energy, which alternatively measures prosodic strength and power, results pointed to an reduction in loudness in repair, which is interesting in itself given that intensity and pitch are commonly found to be correlated. The fact that this is the first study to observe periodic energy in repair makes it worth to keep investigating the effects that are captured by this measurement in terms of prosodic strength and its relationship to marking in repair. I will return to this point in the General Discussion.

The findings described above also tally well with the impressionistic description provided by Goffman (1981, p. 216), who explained that marked repairs, as sampled from English conversations, have their ‘voiced raised and increased tempo’. In our findings, we can associate the ‘voice raising’ to the found increased in pitch, as well as for the described ‘increased tempo’, which we can link to our observed hypospeech. Lastly, regarding Lindblom’s (1990, 1996) H&H theory of informativeness on the hypo-hyper continuum, for the case of self-repair, our findings also support Plug’s view that suggest that speakers may be oriented towards *hypoarticulation*, even if the level of information redundancy in repair may promote *hyperarticulation* (Plug, 2011, p. 296).

HYPOTHESIS B tested whether repairs of incomplete reparanda are more frequently prosodically marked than repairs of complete targets. This expectation derives from prior observations by Nooteboom (2010) and Plug (2016), who reported that repairs with incomplete reparanda (often associated with early detection) tend to show prosodic marking more often than repairs with complete reparanda. In the current data, however, an effect of completeness was found in only one of the three prosodic parameters analysed. Specifically, a small but statistically credible effect was observed in periodic energy, where repairs of incomplete targets exhibited reduced strength, averaging 0.05 units lower than repairs of complete targets. In contrast, pitch and speech rate showed no robust changes as a function of reparandum completeness. Put together, these results indicate that completeness influenced prosodic marking in only a limited and unexpected way. Not only was the effect restricted to a single prosodic parameter, but the direction of the effect also contradicted the hypothesis: rather than displaying enhanced marking, incomplete repairs showed a reduction in prosodic strength. There is therefore no clear evidence in our data to support the idea that repairs resulting from incomplete reparanda are more frequently or more strongly prosodically marked than repairs produced after a fully articulated reparandum. Nor do the results provide support for the broader theoretical prediction derived from Levelt (1989), Levelt et al. (1999), and Hartsuiker and Kolk (2001) that prosodic marking patterns may reflect distinct detection stages in speech monitoring. These models propose two stages of monitoring: one for inner speech (early detection) and one for overt speech (late detection), which might be expected to give rise to different prosodic behaviours by speakers. However, the lack of consistent effects of completeness across the studied prosodic parameters, together with the absence of any bimodal distribution in the yielded prosodic deltas, offers little empirical support for a dual-mode prosodic signalling of detection status in our data. Lastly on HYPOTHESIS B, it is worth noting that the present findings are based on spontaneous occurring repairs, in contrast to studies such as Nooteboom's (2010), which analysed experimentally elicited phonological repairs. While elicited designs offer greater control over repair type and detection timing, they may fail to capture some aspects of spontaneous speech due to additional sources of variability. As shown in Table 15, the distribution of complete and incomplete repairs used in this prosodic analysis reflects the dynamics of naturally produced speech, rather than controlled experimental balance. The absence of consistent prosodic differences based on reparandum completeness observed here may therefore reflect the interactional and variable nature of spontaneous self-repair and highlights the value of studying prosody in naturally occurring speech across different repair types and languages.

HYPOTHESIS C was also based on both (Levelt & Cutler, 1983) and Nooteboom's (2010) findings that early detected errors are more likely to be prosodically marked than late-detected errors. This hypothesis complements Hypothesis B by integrating the role of editing terms as a bridge and a complementary bridge to the detection status. Given the association between completeness and the timing of error detection, I hypothesised that editing terms and prosodic marking might serve partially complementary functions and would therefore be inversely associated. Specifically, I expected that repairs with complete reparanda, typically indicative of *late* detection, would more often be accompanied by editing terms, while repairs with incomplete reparanda, associated with *early* detection, commonly found to be prosodically marked, would show stronger prosodic marking in the absence of editing terms. However, the models revealed no credible evidence of an interaction between editing terms and completeness across any of the three prosodic parameters analysed. While some factors showed independent effects in certain models, combining them in an interaction term did not reveal any systematic change in the prosodic realisation across the models analysing periodic energy, pitch or speech rate. The posterior distributions for the interaction terms were centred around zero, and their 95% credible intervals consistently included zero. This suggests that any interaction effect, if it exists, is likely to be small or highly variable. Nevertheless, it is relevant to acknowledge that the descriptive patterns in the data provide partial support for the predicted trend. Editing terms were generally more frequent in repairs having complete reparanda than in their incomplete counterparts.

Editing terms accompanied 34.2% of repairs classified as having complete reparanda (79 out of 231), compared to only 23.6% classified as incomplete (41 out of 174). Although this difference did not yield a statistically credible interaction in our Bayesian models, the direction of the pattern aligns with the original hypothesis and supports the idea that editing terms are more likely to occur with late-detected repairs (i.e., repairs with complete reparanda). Taken together, these findings suggest that editing terms and prosodic marking do not function as strict alternatives in the signalling of repair, at least not in a mutually exclusive manner. Instead, they may represent partially overlapping resources, deployed independently or in parallel to manage self-repair, and which might depend on contextual constraints. While the absence of a credible interaction limits the strength of this conclusion, the observed descriptive tendencies warrant further investigation using a larger data set and, potentially, more fine-grained discourse distinctions based on the findings reported here. In particular, while the dependent variables in this analysis consisted of continuous acoustic measurements, the predictors were categorical representations of more abstract interactional constructs, such as timing of detection based on reparandum completeness and the categorical distinction between the presence or absence of editing terms. It is therefore possible that the relationship between editing terms and prosodic marking varies in ways that are not fully captured by, for example, the binary categories of completeness and editing terms. Future work might benefit from a more detailed classification of repairs, possibly a taxonomy for editing terms or speaker-specific discourse strategies, to better capture the variability observed in prosodic marking across contexts. Finally, it is important to acknowledge that the relatively small number of repairs including editing terms, particularly in the incomplete category ($n = 41$), may have contributed to the absence of a detectable interaction. Given the natural occurrence of repairs in spontaneous speech, the distribution across categories could not be predetermined, and the resulting sample sizes for certain combinations were limited. While this does not undermine the overall findings, it does suggest that limited statistical power may have played a role and should be considered when interpreting the null interaction result.

Finally, HYPOTHESIS D projected that the repair type on its own would not predict prosodic marking. Results provide no strong evidence that semantic repair categories, namely, appropriateness repairs versus factual and linguistic errors, predict consistent differences in prosodic marking. Across all three prosodic parameters, the effects associated with *Repair Type* are statistically uncertain, with credible intervals overlapping zero in nearly all cases. These findings suggest that the semantic repair type, as defined here, does not play a robust role in shaping the prosodic realization in self-repair. Some weak trends were observed, most notably a tendency for appropriateness repairs to be exhibit less loudness (lower periodic energy) than repairs of linguistic or factual errors. This pattern is in line with the proposed expectation that error repairs might be more strongly marked (Levelt & Cutler, 1983). However, the absence of consistent effects across parameters, and the high degree of posterior uncertainty, prevent firm conclusions from being drawn. Overall, these findings reinforce the perspective presented by Plug (2016) and Plug & Carter (2013) who found semantic distinctions alone did not account for prosodic variation in repair. Rather than reflecting categorical differences in meaning or error type, prosodic marking appears to be shaped more dynamically by contextual, interactional, and speaker-specific factors. This interpretation aligns with Goffman's (1981) view of repair prosody as emergent from the interactional moment, and with Plug's (2016, p. 53) proposal that the role of informativeness in shaping prosody is contextually constrained.

In addition to the overall lack of robust effects for *Repair Type*, further exploration of periodic energy revealed noteworthy speaker-specific variation. While appropriateness repairs showed a general trend toward reduced prosodic strength, this pattern was not consistently observed across other repair types or prosodic parameters. However, analysis of random effects in the periodic energy models showed that speakers differed notably in how they prosodically marked their repairs. While some speakers consistently reduced energy, others tended to enhance it. For example, Speaker 26 produced multiple repairs with clear reductions in mean and minimum periodic energy, whereas Speaker 29 showed consistent increases in prosodic strength across their repair sequences. These individualized marking strategies suggest that, beyond semantic categorization, prosodic realization of repair may be shaped by speaker-specific preferences, possibly reflecting stylistic or interactional factors. Moreover, the observed contrast between stable maximum values and variable mean and minimum values of periodic energy supports the interpretation that these measures index distinct aspects of prosodic effort. I will discuss this interpretation in more depth in the General Discussion. Taken together, these findings reinforce the idea that prosodic marking of repair is not tightly bound to semantic repair type but is instead mediated by contextual and individual factors, a conclusion that aligns with earlier observations on the emergent, situated nature of repair prosody (Goffman, 1981; Plug, 2015a, 2016)

Chapter 5

Repair in Colombian Spanish: Insights and Implications

This chapter discusses the combined implications of the temporal and prosodic analyses of lexical self-repair in Colombian Spanish, with the goal of addressing the overarching research question posed in this thesis, which is assessing the extent to which the findings on the phonetics of lexical self-repair in other languages generalize to the proposed Spanish variety. Chapter 3 and Chapter 4 separately examined the temporal organization and prosodic marking of repair sequences, each through the investigation of a common set of predictors: the semantic repair type, reparandum completeness, and the use of editing terms. The convergence of these independent variables across domains allowed for a structured comparison of their effects, revealing to what extent timing and prosodic behaviour are interrelated or functionally independent. By examining each predictor in turn, this chapter integrates both analyses to evaluate how speakers manage self-repair. Findings indicate that in Spontaneous Colombian Spanish speech, speakers' behaviour responds to flexible, context-sensitive strategies shaped by interactional and cognitive pressures rather than to fixed mappings represented in their acoustic features or temporal realizations.

The chapter is structured as follows: Section 5.1 assesses the limitations of semantic repair categories in predicting timing and prosody. Section 5.2 focuses on the role of reparandum completeness as a predictor of both interruption (i.e., Target-to-cut-off) and resumption (i.e., Cut-off-to-repair) patterns, while it also considers its more selective influence on prosodic realization. Section 5.3 examines the influence and interactional function of editing terms, considering their role in strategic delay of repair. Section 5.4 turns to emergent repair profiles derived from clustering analysis, offering insight into how speakers coordinate repair under varying cognitive and interactional constraints. Section 5.5 reflects on speaker-specific variation and methodological advances in prosodic analysis. Finally, Section 5.6 presents future directions in the study of self-repair, and Section 5.7 concludes with a synthesis of the findings and discusses broader implications for models of speech monitoring and the cross-linguistic study of self-repair.

5.1 The Limits of Semantic Repair Categories in Predicting Timing and Prosody

The semantic classification of repairs, which distinguished between appropriateness and error repairs (including both linguistic and factual errors), did not predict consistent effects across the majority of analyses. While Cutler (1983) and Levelt and Cutler (1983) proposed that error repairs elicit stronger contrastive marking, this pattern did not hold consistently in our spontaneous data. Instead, the influence of *Repair Type* was limited to a small number of specific interactions or weak trends, suggesting that its explanatory power in this data set was restricted to certain repair configurations. For instance, while *Repair Type* alone did not significantly affect Target-to-cut-off or Cut-off-to-repair intervals, editing terms were more frequently observed in appropriateness repairs, which in turn influenced Target-to-cut-off durations. Similarly, *Repair Type* did not predict cluster membership in the timing analysis, indicating that broader repair trajectories are not shaped solely by semantic class. In the prosodic domain, no robust effects of *Repair Type* were found across pitch, speech rate, or periodic energy. A weak tendency was observed for appropriateness repairs to be prosodically weaker, particularly in periodic energy, but this pattern did not show statistical robustness.

Overall, these results suggest that while *Repair Type* may play a role in specific contexts, it does not operate as a consistently strong predictor of repair behaviour in our Spanish spontaneous speech. These findings question the idea that semantic distinctions systematically affect repairs, either temporally or prosodically. Our results support Plug's (2015a) claim that the functional organization of repair is more context sensitive. Appropriateness repairs appear to be managed through lexical strategies, such as editing terms, rather than prosodic emphasis, further reinforcing the idea that semantics alone are insufficient predictors of repair realization. Additional support for this view comes from the cluster analysis discussed in Section 5.4, which discusses distinct temporal organization that cut across appropriateness repairs, highlighting more nuanced repair strategies that speakers deploy in real time.

5.2 Early versus Late Repairs: Temporal Patterns and Prosodic Strategies

Among the predictors examined in this thesis, reparandum completeness was the only one to show a robust effect across both temporal and prosodic domains, although in the latter case, the effect was limited to a single parameter: Periodic energy. This section discusses the findings related to completeness as an individual factor, examining how early versus late interruption patterns influence the organization of repair. Section 5.4 complements this analysis by drawing on cluster-based patterns that highlight how completeness interacts with other timing strategies across the full trajectory of repair.

In line with prior work (e.g., Blackmer & Mitton, 1991; Hartsuiker & Kolk, 2001; Nooteboom & Quené, 2019; Plug & Carter, 2014) incomplete reparanda, which reflect early interruption, were associated with shorter Target-to-cut-off intervals and were frequently followed by shorter Cut-off-to-repair durations. This combination is typically interpreted as evidence of rapid detection and efficient resumption. However, as Nooteboom (2010) argues, such early repairs do not reflect successful prevention of error articulation; they result from internal monitoring detecting the error during the final stages of speech planning, just before articulation begins, but not early enough to fully suppress the onset of the erroneous item. This study contributes to the discussion by showing that such patterns hold not only in elicited phonological repairs, which have been a primary focus research, but are also observable in spontaneous lexical repairs, which are complex in both planning and communicative function. Lexical repairs typically involve semantically informative material and unfold in interactionally rich contexts, where the speaker needs to consider the lexical units meaning, appropriateness and social factors, most probably, under time pressure. The fact that spontaneous lexical repairs in this dataset exhibit similarly short interruption and resumption intervals as those reported for phonological repairs in experimental studies indicates that comparable timing patterns can arise even in semantically and interactionally complex contexts. This observation also directly addresses the question raised by Nooteboom and Quené (2017) and Quené and Nooteboom (2024), namely that findings from phonological error repairs may not generalize straightforwardly to spontaneous or lexical repairs. As they point out, the two types of repairs likely involve different underlying processes, and empirical confirmation in naturalistic data has remained scarce. As they note (Quené & Nooteboom, 2024) the ~460 ms split observed in their elicited data may underestimate the true range of variation due to task constraints. This study helps address that limitation: as will be discussed in Section 5.4, the clustering analysis reveals a wider array of timing profiles, confirming that early and late detection dynamics also emerge in spontaneous, semantically driven repair.

Despite these temporal findings, prosodic realization did not align with theoretical expectations. Early accounts on the matter (e.g., Levelt & Cutler, 1983; Plug, 2014) suggest that early-detected repairs, particularly those correcting salient errors, should be prosodically enhanced through pitch, duration, or intensity. Yet, in this study, incomplete reparanda were not associated with elevated pitch or speech rate. Instead, they showed a statistically credible reduction in periodic energy, while other prosodic parameters remained unaffected. This challenges the assumption that early detection triggers prosodic emphasis and suggests that speakers may instead adopt a softly managed response and a fluency-oriented approach. The reduction in periodic energy, along with stable speech rate and pitch, may reflect a hesitant or *low-commitment* prosodic configuration in repair, where the speakers' goal is not to contrast the correction with the error, but simply to move forward in the interaction.

Taken together, these findings reveal a division between the temporal organization of repair and its prosodic realization. While completeness robustly predicts interruption and resumption intervals, prosodic marking appears less predictable, and more dependent on speaker-specific or interactional considerations. The fact that periodic energy and not pitch nor speech rate captured this effect also highlights the value of using multi-parametric prosodic measures to access different layers of speaker behaviour. These observations support a view of prosodic marking not as a mechanical output of monitoring, but as a flexible, context-sensitive resource that speakers can deploy or withhold based on moment-to-moment interactional demands. This interpretation aligns with the interactional models of Goffman (1981) and Plug (2015a) who conceptualize repair as an emergent practice shaped by discourse contingencies, speaker stance, and pragmatic goals.

This pattern suggests that attentional engagement during overt speech monitoring in spontaneous language is not necessarily lower than during inner speech monitoring; rather, both stages of monitoring appear equally engaged, but are modulated by the objective of the interaction and the resulting type of speech. Specifically, inner monitoring may be more active in experimental settings, while overt monitoring appears to be more crucial in spontaneous speech. Nooteboom argues that this pattern is possible because, in elicited errors, speakers do not need to monitor for lexical conflicts, whereas monitoring for lexical appropriateness is a key aspect of spontaneous communication (Nooteboom, 2010, pp. 231-232).

5.3 Editing Terms as Lexical Strategies for Managing Repair Timing

Editing expressions have long been recognized as a functional and integral component of self-repair, particularly in their role of signalling trouble and structuring the transition between interruption and resumption (Levelt, 1989; Nooteboom, 2005b, 2010; Plug, 2016). Although their distribution and function may vary across repair types, authors working within both monitoring-based and interactional frameworks have emphasized their importance in helping speakers deal with disruption. The findings of this study, based on spontaneous lexical self-repairs in Colombian Spanish, offer empirical support for these claims, demonstrating that editing expressions are not only present in unscripted speech, but help implementing pragmatically motivated strategies in repair organization. In this study, they emerged as strong predictors of repair timing, but not of prosodic realization. Their presence was systematically associated with longer Cut-off-to-repair intervals and more importantly, with longer Target-to-cut-off durations, as well. These patterns reflect what previous authors (Seyfeddinipur et al., 2008; Nooteboom & Quené, 2019) have described as a ‘repair delay’ strategy. Here, the presence of editing terms appears go in line with a delayed cut-off, suggesting that speakers are not interrupting early, but rather holding off interruption while they assess or reconsider their utterance. This suggests that editing terms serve strategically as they can help the speaker signal trouble to the listener while continuing to hold the floor, as explained by Seyfeddinipur et al. (2008, p. 841). Interestingly, this pattern was primarily observed in appropriateness repairs, cases where the trouble source may not be strictly erroneous, but contextually or socially sensitive. In such cases, speakers may delay interruption to evaluate the adequacy of their utterance, and then use an editing term as a pragmatic bridge between internal decision-making and overt reformulation. This would result in a temporal profile characterized by early detection, late interruption, and early resumption, a configuration that matches our observed Late-Early cluster and which provides further support to the 'strategic delay' proposed by Seyfeddinipur et al. (2008) .

At the same time, editing terms were not associated with consistent changes in prosody. There was no reliable effect on pitch, periodic energy, or speech rate, and no credible interaction between editing terms and completeness in prosodic realization. This asymmetry, strong effects on timing but no systematic prosodic features, supports the view that editing terms operate as a lexical strategy rather than a prosodic one. Findings suggest that they are used in cases where the speakers need extra time, signal trouble, or bridge moments of uncertainty, but not necessarily to enhance the salience of the correction. Importantly, these findings align with Nooteboom and Quené's (2019, p. 81) proposal that editing expressions serve as a delaying strategy in spontaneous speech, especially when overt monitoring is in use. This builds on Nooteboom's (2010) broader view that once a speech error becomes public, the speaker is no longer under pressure to prevent it from surfacing, and therefore, has less need to signal the repair with additional prosodic emphasis. In such contexts, editing terms can help speakers lexically handle the presence of a repair while allowing them to maintain fluency.

In addition, these findings offer further support for Plug's (2015a) view that the functional organization of repair is context-sensitive and pragmatically motivated. While Levelt and Cutler (1983) proposed that errors elicit stronger prosodic marking than appropriateness repairs, that pattern did not hold in our data. Instead, appropriateness repairs were often managed lexically through editing terms, rather than prosodically. This suggests that speakers flexibly choose between lexical and prosodic strategies based on situational demands.

In sum, editing terms emerge in this analysis as a central resource for implementing repair strategies without disrupting flow. This division of labour observed in our spontaneous data, with editing terms supporting lexical repair strategies and prosody remaining relatively unmarked, highlights the need to understand repair not merely as a structural process, but as an interactionally rooted, real-time practice.

5.4 Emergent Repair Profiles: Evidence from Timing Clusters

Beyond the individual contributions of predictors such as completeness and editing terms, the joint distribution of Target-to-cut-off and Cut-off-to-repair intervals revealed a more nuanced picture of how speakers coordinate self-repair. Through k-means clustering, three distinct repair profiles emerged: Early-Early, Early-Late, and Late-Early. These clusters reflect combined timing dynamics that offer insight into speakers' strategies, cognitive load, and fluency management, aspects that were not fully captured by the individual predictors alone.

The Early-Early cluster, characterized by short intervals on both sides, reflects relatively early detection and rapid resolution: the speaker interrupts early and resumes quickly. This can be seen as an example of monitoring-based efficiency, likely corresponding to detection in inner speech. In contrast, the Early-Late cluster suggests a different strategy: interruption occurs early, but repair initiation is delayed. This pattern is robustly associated with the presence of editing terms, which appear to function as a lexical tool for managing hesitation and planning. These repairs also tend to cluster with appropriateness repairs, which require more contextual elaboration than errors. The Late-Early cluster, on the other hand, challenges the expectations from dual-loop models (e.g., Levelt, 1989) as it features late interruption but rapid repair. This may indicate that planning was already underway before the cut-off, or that the delay was a strategic choice, not simply a function of detection latency.

Crucially, the absence of a robust Late-Late cluster reinforces the idea that in spontaneous speech, extended disruption is interactionally costly. Speakers appear reluctant to delay interruptions, supporting the idea that repair strategies are selected in real-time to balance efficiency, clarity, and fluency. This pattern aligns closely with the account proposed by Seyfeddinipur et al. (2008) who argue that speakers can adjust the timing of interruption and resumption strategically, even after early error detection. The clustering patterns observed here, particularly the presence of Late-Early and Early-Late configurations, offer empirical support for this view, suggesting that speakers manage repair timing flexibly, rather than being bound to fixed response profiles following detection.

These findings offer an important complement to the predictor-based analyses discussed above. From them we can hypothesize that while individual predictors highlight factors influencing repair structure, the clustering analysis captures how these factors combine dynamically in real-time. This perspective is broadly consistent with interactional accounts of repair (e.g., Goffman, 1981; Plug, 2015a), which emphasize that repair behaviour is situated, variable, and responsive to the demands of the moment. Although this study did not examine interactional sequences in detail, these frameworks remain relevant for future work. In particular, further research might explore how these timing profiles relate to speaker identity, prosodic style, and epistemic positioning, especially in contexts where confidence and hesitation fluctuate. Integrating such approaches would build on the present findings and strengthen the link between timing strategies and the broader interactional organization of repair.

5.5 Prosodic Marking in Spontaneous Repair: Speaker-Specific Strategies and Methodological Advances

One of the most distinctive findings of the prosodic analysis was the presence of speaker-specific variation in periodic energy, suggesting that individual speakers manage the prosodic realization of repair in different ways. While some speakers, such as Speaker 26, consistently produced repairs with reduced mean and minimum periodic energy, potentially reflecting a strategy of minimizing prosodic prominence, others, like Speaker 29, showed a consistent enhancement of prosodic strength across repairs. These patterns support the idea that prosodic marking is not governed uniquely by structural factors like the type of repair, but it may also be modelled by stylistic or individual preferences and interactional constraints.

Crucially, the above-described observations were made possible through the inclusion of periodic energy as a phonetic correlate of prosodic prominence. Unlike pitch or speech rate, which did not show robust effects in relation to *Repair Type* or *Completeness*, periodic energy appears to have captured subtle variations in prosodic strength that aligned with both speaker-level strategies and the presence of epistemically motivated hesitation. Following recent models of prominence that emphasize the interplay between F0 and periodic energy (Albert, 2023, pp.

146-147), this thesis expands the methodological value of incorporating periodic energy into prosodic research by means of the ProPer workflow, which enabled a level of sensitivity to gradience and local variation that is worth further exploring. To our knowledge, this is the first application of periodic energy to the study of self-repair prosody in spontaneous speech. The findings reported here offer preliminary support for its potential as a diagnostic of speakers' stance and repair strategies. Future studies are encouraged to test the generalizability of these findings across languages, corpora, and prosodic environments.

In addition to periodic energy, this study also implemented a dynamic measurement of pitch ($\Delta F0$) using the ProPer workflow, based on the steepest syllable-to-syllable increase in F0 within each repair sequence. This measure, which complements more traditional peak-based F0 values, reflects a growing interest in capturing pitch movement across units of interests rather than relying solely on static targets. In this dataset, results from $\Delta F0$ were largely consistent with those obtained from conventional peak F0 comparisons, suggesting that both methods capture similar prosodic contrasts in repair. However, the successful implementation of $\Delta F0$ points to a promising direction for future work, especially in studies that aim to track fine-grained, perception-relevant fluctuations in pitch across the course of an utterance. The inclusion of this dynamic measure, along with periodic energy, highlights the analytical power of ProPer as a tool for advancing prosodic research in spontaneous speech.

ProPer also offers practical advantages for researchers working with spontaneous speech. While this thesis provided manually transcribed material both at both the syllabic and segmental levels, ProPer can operate fully autonomously using a signal-based segmentation algorithm that detects syllabic boundaries based on fluctuations in periodic energy, an approach based in principles of acoustic salience that correlate with syllabic nuclei (Albert, 2023, pp. 146-147). In informal comparisons conducted during this project, the automatic segmentations generated by ProPer aligned closely with provided annotations, supporting its applicability even when full segmental annotation is unavailable. Given the substantial effort required to transcribe and segment spontaneous data, this functionality may expand the accessibility of prosodic analysis in under-resourced languages or large datasets.

Taken together, the patterns observed in periodic energy suggest that further work to explore how speakers signal hesitation, doubt, or epistemic caution in prosodic terms would be welcome. A more detailed discourse-pragmatic treatment would allow future research to link prosodic behaviour to specific sequences and interactional functions in greater detail.

5.6 The Future Research Value of the UCSI Corpus

Although the analyses presented in this thesis focused primarily on the temporal organization and prosodic features of self-initiated lexical repairs, the UCSI Corpus offers broader potential for future research. As outlined in Chapter 2, the corpus was designed to meet several key criteria, capturing spontaneous, unscripted speech; achieving dialectal representativeness; providing substantial speaker-specific data; and ensuring a high density of repair sequences. These features collectively establish the corpus as a robust, reusable dataset for examining diverse phenomena in Colombian Spanish.

One of the most significant contributions of the UCSI Corpus lies in its richness of repair data. With over 1,000 repair sequences, 923 of which are self-initiated, self-repairs, the corpus enables not only focused investigation of dominant patterns but also exploratory work on less frequent repair types, such as other-initiated repair. These less common subtypes, though not analysed in the present thesis, offer valuable opportunities for future research on interactional features, turn-taking dynamics, and social indexing in repair. At the same time, its design and scope make it a valuable resource for the study of Colombian Spanish more broadly, offering opportunities for research on regional variation, discourse organization, prosody, and interactional dynamics beyond repair.

Moreover, the dialectal and sociolinguistic richness of the corpus positions it well for cross-dialectal comparisons of repair practices. Future studies might examine whether the temporal and prosodic features of repair differ systematically across regional varieties or sociolinguistic profiles. In addition, the corpus supports investigation of repair in contexts not addressed here, including phonological and syntactic repairs, which may reveal further complexity in planning and monitoring mechanisms.

Taken together, the UCSI Corpus not only supports the empirical findings of this thesis but also stands as a valuable resource for advancing the study of variation and interaction in Spanish spontaneous speech.

5.7 Future Directions: Toward Pragmatic and Individualized Models of Repair

The findings discussed within this chapter contribute to the broader goal of evaluating whether the phonetic patterns observed in self-repair across other languages extend to Colombian Spanish, a language and variety largely underrepresented in both phonetic and interactional research on repair. This thesis confirms several established patterns, such as the link between early detection and faster repairs, suggesting that core aspects of the temporal organization of repair generalize across languages. However, the results also reveal points of divergence, particularly in the prosodic realization of repair, where variation across speakers and repair subtypes suggests a more context-dependent use of repair marking strategies than previously assumed.

These observations open several directions for future research. First, the role of speaker-specific prosodic strategies in repair could benefit from further research. While previous work has examined repair marking as a function of detection stage or the semantic type of repair (e.g., Levelt & Cutler, 1983; Nooteboom, 2005), This dissertation has shown that such factors do not fully account for how prosodic marking unfolds in spontaneous repair. Also, comparatively, there is little knowledge about how individual speakers vary in their use of prosody across repairs and how these tendencies correlate with social, stylistic, or interactional roles. Research has pointed to the value of examining prosodic cues as discourse constraints as faced by speakers (Plug, 2015a), or the epistemic position by participants in conversations (Sidnell & Barnes, 2013) but few studies have systematically linked these insights to phonetic variability across speakers. The UCSI corpus, developed for this thesis, is designed to support such analyses. It includes sociolinguistic variables and is structured around different task-based interactions. There, we can see speakers positioned in varied interactional and epistemic configurations. Although the present study focused on lexical self-initiated, self-repair, the corpus also contains data on other repair types such as phonological repairs, and other configurations on who initiates and performs repair (e.g., self-initiated, other-repair, other-initiated, self-repair or other-initiated, other-repair) offering a foundation for future research on repair across extra configurations and modalities, both within Spanish and cross-linguistically.

Second, more detailed analysis of editing terms and repair taxonomies would support the development of more precise models of how lexical scaffolding and prosodic marking function in combination. Much of the existing typology of editing terms is drawn from English (e.g., Clark & Fox Tree, 2002); and even Plugs (2016) work with corpus spontaneous data, which motivated many aspects of this research, leaves room for expansion into other languages. For Spanish, there is still no work on how speakers use editing expressions such as *eee* 'uh', or *pues* 'well' all found in this thesis' sampled repairs, and even less we know on how these expressions differ by *Repair Type* or interactional context. Future research could combine fine-grained syntactic and pragmatic annotation of editing terms with acoustic analysis, to explore when and how these expressions signal planning trouble, social sensitivity, or discourse management. In particular, a more detailed taxonomy of Spanish repair subtypes, possible drawing from previous work in English, and including recent conversation-analytic research, would support cross-linguistic comparisons of how trouble sources are identified and resolved in spontaneous talk.

Finally, this thesis highlights the methodological inclusion of periodic energy as a prosodic measure. Traditional prosodic research has often focused on pitch (F0) and intensity and the speed of speech by observing articulatory or speech rate. The use of periodic energy, as implemented here through the ProPer workflow (Albert et al., 2024), revealed meaningful differences in repair marking in addition to the information revealed by pitch and speech rate. This suggests that periodic energy may be useful in capturing other nuanced prosodic patterns associated with speakers' positioning around the repair setting. However, further research is needed to validate this method in other speech contexts, including, for example, controlled experimental data, cross-linguistic corpora, and discourse environments beyond self-repair.

These research directions point towards a more speaker-sensitive, interactionally grounded account of repair, in which prosody is not a rigid correlate of structural position, but a resource deployed flexibly in response to planning needs, social context, and epistemic factors. Expanding this line of work across languages can help inform our current phonetic models of prosodic prominence and the understanding we hold on to interactional linguistic behaviour, contributing to a richer account of how speakers manage disruption in everyday talk.

5.7 General Conclusion: Integrating Temporal and Prosodic Perspectives on Lexical Self-Repair

This chapter has evaluated how the phonetic patterns of lexical self-repair observed in other languages extend to spontaneous Colombian Spanish, by integrating findings from temporal and prosodic analyses. The results discussed in this chapter demonstrate that key aspects of temporal structure, such as the relationship between early detection, shorter interruption intervals, and faster resumption, are consistent with findings from other languages, including those based on elicited speech. These patterns provide cross-linguistic support for models of inner speech monitoring and suggest that speakers, across languages, share general mechanisms for recognizing and responding to lexical trouble.

At the same time, the prosodic analysis revealed a context-sensitive and speaker-specific picture. Unlike earlier accounts which proposed consistent prosodic marking for early or semantically weighty repairs, the findings here suggest that prosodic marking is not systematically governed by *Repair Type* or the detection stage, as observed through completeness. Instead, speakers appear to adopt a repertoire of strategies, using prosody flexibly and sometimes opting instead for lexical tools, such as editing terms, to manage planning pressure or mark trouble.

One methodological contribution of this study was the implementation of periodic energy as a prosodic measure. While pitch and speech rate showed limited effects, periodic energy revealed speaker-specific variation that may reflect subtle forms of hesitation, uncertainty, or stylistic choice. Its successful application in this study suggests its potential value in future analyses of spontaneous speech.

Taken together, the findings from this thesis support a view of self-repair as a flexible, multi-layered process, shaped by both cognitive and interactional demands. While temporal patterns reflect underlying monitoring mechanisms that appear to generalize across languages, prosodic marking is best understood as an emergent, speaker-driven practice, sensitive to context, fluency, and communicative goals.

Chapter 6

Conclusions

This thesis examined the timing and prosodic realization of self-initiated lexical repair in unscripted Colombian Spanish, drawing on original spontaneous data. It explores both Target-to-cut-off and Cut-off-to-repair, as relevant timing intervals for describing the temporal organization of repair, and a series of prosodic parameters (periodic energy, pitch, and speech rate) to examine prominence. The findings from the two experimental chapters converge on several critical insights.

First, regarding temporal organization, the results confirm that the timing of repairs is highly variable and context-sensitive. The wide distributions observed for both Target-to-cut-off and Cut-off-to-repair intervals align with prior reports of early and late detection mechanisms in self-monitoring (Hartsuiker & Kolk, 2001; Levelt, 1989). Clustering analysis revealed three dominant temporal configurations, Early-Early, Early-Late, and Late-Early, indicating that interruption can be subject to strategic delay, depending on the speaker's processing constraints and interactional goals, which is in line with the idea that speakers can delay interruption to maximize fluency (Seyfeddinipur et al., 2008). Interestingly, the anticipated Late-Late configuration, following Nooteboom's (2010) argument that relates overt speech with prolonged Cut-off-to-repair intervals, was not robustly observed, suggesting that late detection does not necessarily entail prolonged delays throughout the repair sequence. Instead, the data point toward a pragmatic use of delay: either at the point of interruption (Late-Early) or at the repair onset (Early-Late) which appears to be managed through editing terms. In addition, the temporal organizations observed provide evidence not only to the flexible nature of repair timing, but also to a preference for maintaining fluency in spontaneous interaction. Speakers appear to avoid prolonged disruptions, favouring strategies (such as editing terms or brief delays at one stage only) that minimize perceptible breakdowns. Rather than late detection invariably leading to late repair, the findings suggest that speakers may actively modulate when and how they pause, deploying prosodic and lexical cues to manage trouble while preserving the overall flow of talk.

Second, with respect to prosodic marking, speakers tend to mark repairs by producing them with faster articulation rates and higher pitch, in line with Plug's (2014) proposal that prosodic marking reflects a trade-off between emphasis and efficiency. However, periodic energy, a novel measure in this thesis, was found to decrease in repairs, a result which diverges from the more commonly reported increase in intensity. This suggests that periodic energy captures a distinct aspect of prosodic strength and may offer a more nuanced index of prosodic modulation in repair. Notably, reparandum completeness and semantic repair type did not robustly predict pitch or tempo, and only a limited, unexpected effect of *Completeness* emerged for periodic energy.

Third, the presence of editing terms was found to significantly influence timing and interact with semantic type, particularly in appropriateness repairs. Editing terms were associated with both delayed interruptions and faster repair onsets, underscoring their role as both cognitive and discourse resources.

Finally, across both analyses, no strong support was found for semantic repair type (i.e., appropriateness versus factual and linguistic repairs) as a consistent predictor of timing or prosodic realization. While some weak tendencies emerged, such as appropriateness repairs being prosodically weaker, these effects were inconsistent and highly speaker-dependent. Indeed, speaker-specific marking behaviours were a recurrent theme in the case of periodic energy, highlighting the importance of continuing to research individual and contextual variability in repair behaviour.

Taken together, these findings challenge the notion of prosodic marking as a uniformly speaker, or listener-oriented signal. Instead, they support a view of self-repair as a flexibly deployed, context-sensitive behaviour, in which prosodic and temporal cues are modulated by detection timing, cognitive effort, and speaker style. By incorporating measures like periodic energy and by analysing spontaneous speech, this thesis contributes to new empirical and theoretical grounding to models of self-monitoring and repair, while also proposing methodological innovations for studying spontaneous speech production.

List of References

Adobe Inc. (2019). *Audition*. In (Version 22.0.096) https://www.adobe.com/products/audition.html

Albert, A. (2023). *A model of sonority based on pitch intelligibility*. Language Science Press.

Albert, A., Cangemi, F., & Grice, M. (2018). *Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration*. https://doi.org/10.21437/SpeechProsody.2018-162

Albert, A., Cangemi, F., & Grice, M. (2019). *Can you draw me a question?* https://doi.org/10.13140/RG.2.2.15700.14729

Albert, A., Cangemi, F., Grice, M., & Ellison, T. M. (2024). *ProPer: PROsodic analysis with PERiodic energy*. Retrieved August 2023, from https://doi.org/10.17605/OSF.IO/28EA5

Albert, A., & Nicenboim, B. (2022). Modeling Sonority in Terms of Pitch Intelligibility With the Nucleus Attraction Principle. *Cognitive Science*, *46*(7), e13161.

Baars, B. J., & Motley, M. T. (1974). *Spoonerisms: Experimental elicitation of human speech errors*. American Psycholog. Ass., Journal Suppl. Abstract Service.

Baker, R., & Hazan, V. (2011). DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior research methods*, *43*(3), 761-770. https://doi.org/10.3758/s13428-011-0075-y

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1 - 48. https://doi.org/10.18637/jss.v067.i01

Bernal-Chávez, J. A., & Díaz-Romero, C. E. (2017). Caracterización panorámica del español hablado en Colombia: fonología y gramática. *Cuadernos de Lingüística Hispánica*(29), 19-37. https://doi.org/10.19053/0121053x.n29.2017.5845

Blackmer, E. R., & Mitton, J. L. (1991). Theories of monitoring and the timing of repairs in spontaneous speech. *Cognition*, *39*(3), 173-194. https://doi.org/10.1016/0010-0277(91)90052-6

Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer [Computer program]*. In (Version 6.1.56, retrieved 3 November 2021) http://www.praat.org/

Bredart, S. (1991). WORD INTERRUPTION IN SELF-REPAIRING. *Journal of psycholinguistic research*, *20*(2), 123-138. https://doi.org/10.1007/BF01067879

Brown, E. L., & Brown, E. K. (2012). Syllable-Final and Syllabel-Initial/S/Reduction in Cali, Colombia: One Variable or Two? *Syllable-Final and Syllabel-Initial/S/Reduction in Cali, Colombia: One Variable or Two?*, 89-106.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1 - 28. https://doi.org/10.18637/jss.v080.i01

Cangemi, F. (2015). *mausmooth*. In http://ifl.phil-fak.uni-koeln.de/fcangemi.html

Carter, P., & Plug, L. (2011). Prosodic characteristics of interrupted versus completed speech error repairs in spontaneous Dutch speech.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111. https://doi.org/10.1016/S0010-0277(02)00017-3

Coloma, G. (2016). Una versión alternativa de “El viento norte y el sol” en español. *Revista de Investigación Lingüística*, *18*(0), 191-212. https://revistas.um.es/ril/article/view/248061

Coloma, G. (2018). Argentine Spanish. *Journal of the International Phonetic Association*, *48*(2), 243-250. https://doi.org/10.1017/S0025100317000275

Couper-Kuhlen, E., & Selting, M. (2017). *Interactional linguistics : study language in social interaction*. Cambridge University Press.

Cutler, A. (1983). Speakers' Conceptions of the Function of Prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements* (pp. 79-91). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-69103-4_7

de Cheveigné, A. (2005). Pitch Perception Models. In C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper (Eds.), *Pitch: Neural Coding and Perception* (pp. 169-233). Springer New York. https://doi.org/10.1007/0-387-28958-5_6

Departamento Administrativo Nacional de Estadística (DANE). (2025). *¿Cuántos somos?* https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/cuantos-somos

Departamento Administrativo Nacional de Estadísticas (DANE). (2015). *Metodología de estratificación socioeconómica urbana para servicios públicos domiciliarios: Manual de realización (Versión 2.0)*.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., Drew, P., Floyd, S., Gipper, S., Gísladóttir, R., Hoymann, G., Kendrick, K. H., Levinson, S. C., Magyari, L., Manrique, E., Rossi, G., San Roque, L., & Torreira, F. (2013). Huh? What? – a first survey in twenty-one languages. In G. Raymond, J. Sidnell, & M. Hayashi (Eds.), *Conversational Repair and Human Understanding* (pp. 343-380). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511757464.012

Ernestus, M. (2000). Voice Assimilation and Segment Reduction in Casual Dutch.

Espejo Olaya, M. B. (2016). Alofonía de /s/ en Colombia. *Lingüística y Literatura*, *37*(69), 99-117. https://doi.org/10.17533/udea.lyl.n69a04

Figueroa, M. A., García, D. A., & Salamanca, G. F. (2019). DiapixSp: adaptación al español y aplicación piloto de una herramienta de elicitación de habla espontánea y colaborativa. *Estudios de fonética experimental*, 257-288.

File-Muriel, R. J. (2009). The Role of Lexical Frequency in the Weakening of Syllable-Final Lexical /s/ in the Spanish of Barranquilla, Colombia. *Hispania*, *92*(2), 348-360.

Flórez, L. (1961). El Atlas Lingüístico-Etnográfico de Colombia, ALEC. Nota informativa. *Thesaurus: Boletín del Instituto Caro y Cuervo*, *XVI*(1), 77-125.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. CRC Press.

Goffman, E. (1981). Radio Talk. A Study Of The Way Of Our Errors. In: Forms of Talk. Pennsylvania: Pennsylvania University Press.

Goldman, J.-P. (2011). EasyAlign: An Automatic Phonetic Alignment Tool Under Praat. INTERSPEECH,

Goldman, J.-P., & Schwab, S. (2014). Easyalign Spanish: An (Semi-)Automatic Segmentation Tool Under Praat. https://doi.org/10.5167/uzh-107124

Hardy, A. (1994). An examination of procedures for determining the number of clusters in a data set. New Approaches in Classification and Data Analysis, Berlin, Heidelberg.

Hartsuiker, R., Kolk, H., & Martensen, H. (2005). The division of labor between internal and external speech monitoring. In R. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phonological encoding and monitoring in normal and pathological speech* (pp. 187 - 205). Psychology Press.

Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error Monitoring in Speech Production: A Computational Test of the Perceptual Loop Theory [Article]. *Cognitive Psychology*, *42*(2), 113-157. https://doi.org/10.1006/cogp.2000.0744

Hepburn, A., & Bolden, G. B. (2017). *Transcribing for social research*. SAGE Publications.

Instituto Caro y Cuervo. (1981). *Atlas Lingüístico-Etnográfico de Colombia, ALEC*. Instituto Caro y Cuervo.

Instituto Cervantes. (2019). *El español: una lengua viva. Informe 2019*. https://cvc.cervantes.es/lengua/anuario/anuario_19/informes_ic/p01.htm

Instituto Geográfico Agustín Codazzi [IGAC]. (2020). *Resolución No. 471 de 2020*. https://www.igac.gov.co/transparencia-y-acceso-a-la-informacion-publica/normograma/resolucion-no-471-de-2020

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011

James, D. (1972). Some Aspects of the Syntax and Semantics of Interjections. In P. L. Peranteau, JN; Phares, GC. (Ed.), *Papers from the Eighth Regional Meeting of the Chicago Linguistic Society* (pp. 162-172). Chicago Linguistic Society.

Kapatsinski, V. (2010). Frequency of Use Leads to Automaticity of Production: Evidence from Repair in Conversation. *Language and speech*, *53*(1), 71-105. https://doi.org/10.1177/0023830909351220

Kassambara, A., & Mundt, F. (2020). factoextra: Extract and visualize the results of multivariate data analyses (R package version 1.0.7).

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

Ladefoged, P. (2017). *Elements of Acoustic Phonetics*. The University of Chicago Press.

Lancheros Redondo, H. F. (2018). Los indigenismos léxicos en las variedades diatópicas del español colombiano. *Forma y Función*, *31*(2), 9-29. https://doi.org/10.15446/fyf.v31n2.74652

Lemoine, N. P. (2019). Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos*, *128*(7), 912-928. https://doi.org/10.1111/oik.05985

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41-104.

Levelt, W. J. M. (1989). Self-Monitoring and Self-Repair. In. MIT Press.

Levelt, W. J. M., & Cutler, A. (1983). Prosodic Marking in Speech Repair. *Journal of semantics (Nijmegen)*, *2*(2), 205-218. https://doi.org/10.1093/semant/2.2.205

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and brain sciences*, *22*(1), 1-38. https://doi.org/10.1017/S0140525X99001776

Lindblom, B. (1990). Explaining Phonetic Variation: A Sketch of the H&H Theory.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *The Journal of the Acoustical Society of America*, *99*, 1683-1692. https://doi.org/10.1121/1.414691

Malik, A., & Tuckfield, B. (2019). *Applied Unsupervised Learning with R : Uncover hidden relationships and patterns with k-means clustering, hierarchical clustering, and PCA*. Packt Publishing.

Markič, J. (2017). Phonetic and phonological aspects of Colombian Spanish. *Linguistica (Ljubljana)*, *57*(1), 185-196. https://doi.org/10.4312/linguistica.57.1.185-196

Martínez-Celdrán, E., Fernández-Planas, A. M., & Carrera-Sabaté, J. (2003). Castilian Spanish. *Journal of the International Phonetic Association*, *33*(2), 255-259. https://doi.org/10.1017/S0025100303001373

McElreath, R. (2020). *Statistical rethinking : a Bayesian course with examples in R and Stan* (Second edition. ed.). CRC Press, Taylor & Francis Group.

Monroy, R., & Hernández-Campoy, J. M. (2015). Murcian Spanish. *Journal of the International Phonetic Association*, *45*(2), 229-240. https://doi.org/10.1017/S0025100314000231

Montes, J. J. (1982). El español de Colombia. Propuesta de clasificación dialectal. *Thesaurus: Boletín del Instituto Caro y Cuervo*, *XXXVII*(1), 23-92.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational Ideas-Part II. *Language and linguistics compass*, *10*(11), 591-613. https://doi.org/10.1111/lnc3.12207

Nooteboom, S. (1980). Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. *Errors in Liguistic Performance: Slips of the Tongue, Ear, Pen and Hand*.

Nooteboom, S. (2010). Monitoring for speech errors has different functions in inner and overt speech. In M. Everaert, Lentz, T, Mulder, H, Nilsen, O (Ed.), *The Linguistic Enterprise: From Knowledge of Language to Knowledge in Linguistics* (pp. 231-234). John Benjamins. https://doi.org/10.1075/la.150.09noo

Nooteboom, S., & Quené, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory and Language*, *58*(3), 837-861. https://doi.org/https://doi.org/10.1016/j.jml.2007.05.003

Nooteboom, S., & Quené, H. (2014). Do speakers try to distract attention from their speech errors? The prosody of self-repairs. In (pp. 203-217). John Benjamins Publishing Company. https://doi.org/10.1075/z.189.17noo

Nooteboom, S. G. (2005a). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech communication*, *47*(1), 43-58. https://doi.org/10.1016/j.specom.2005.02.003

Nooteboom, S. G. (2005b). Listening to oneself: Monitoring speech production. In *Phonological encoding and monitoring in normal and pathological speech.* (pp. 167-186). Psychology Press.

Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of memory and language*, *95*, 19-35. https://doi.org/10.1016/j.jml.2017.01.007

Nooteboom, S. G., & Quené, H. (2019). Temporal aspects of self-monitoring for speech errors. *Journal of memory and language*, *105*, 43-59. https://doi.org/10.1016/j.jml.2018.11.002

Oomen, C. C. E., & Postma, A. (2001). Effects of Time Pressure on Mechanisms of Speech Production and Self-Monitoring. *Journal of psycholinguistic research*, *30*(2), 163-184. https://doi.org/10.1023/A:1010377828778

Oomen, C. C. E., & Postma, A. (2002). Limitations in processing resources and speech monitoring. *Language and cognitive processes*, *17*(2), 163-184. https://doi.org/10.1080/01690960143000010

Oostdijk, N. (2002). The design of the Spoken Dutch Corpus. In. https://doi.org/10.1163/9789004334113_008

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, *106*(1), 528-537. https://doi.org/https://doi.org/10.1016/j.cognition.2007.02.006

Orduz Navarrete, Y. (2013). La pérdida de distinción fonológica /ɉ/ - /ʎ/ en el habla bogotana. *Folios (Bogotá, Colombia)*(37), 51-62.

Oxenham, A. J. (2012). Pitch perception. *The Journal of neuroscience*, *32*(39), 13335-13338. https://doi.org/10.1523/JNEUROSCI.3815-12.2012

Plug, L. (2010). Pragmatic constraints in Usage-based Phonology, with reference to some Dutch phrases. *Journal of pragmatics*, *42*(7), 2014-2035. https://doi.org/10.1016/j.pragma.2009.12.001

Plug, L. (2011). Phonetic reduction and informational redundancy in self-initiated self-repair in Dutch. *Journal of phonetics*, *39*(3), 289-297. https://doi.org/10.1016/j.wocn.2010.08.001

Plug, L. (2014). On (or not on) the 'upgrading-downgrading continuum': the case of 'prosodic marking' in self-repair. *Prosody and Phonetics in Interaction*, 70-86.

Plug, L. (2015a). Discourse constraints on prosodic marking in lexical replacement repair. *Journal of pragmatics*, *87*, 80-104. https://doi.org/10.1016/j.pragma.2015.07.005

Plug, L. (2015b). Prosodic marking and predictability in lexical self-repair.

Plug, L. (2016). Informativeness, Timing and Tempo in Lexical Self-Repair. *Language and speech*, *59*(4), 516-543. https://doi.org/10.1177/0023830915618427

Plug, L., & Carter, P. (2013). Prosodic Marking, Pitch and Intensity in Spontaneous Lexical Self-Repair in Dutch. *Phonetica*. https://doi.org/10.1159/000355512

Plug, L., & Carter, P. (2014). Timing and tempo in spontaneous phonological error repair. *Journal of phonetics*, *45*(Jul), 52-63.

Plug, L., Lennon, R., & Smith, R. (2023). Testing for canonical form orientation in speech tempo perception. *Quarterly Journal of Experimental Psychology*, *77*(7), 1443-1461. https://doi.org/10.1177/17470218231198344

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition*, *77*(2), 97-132. https://doi.org/10.1016/S0010-0277(00)00090-1

Quené, H., & Nooteboom, S. G. (2024). Some properties of mental speech preparation as revealed by self-monitoring [Article]. *Speech Communication*, *158*, Article 103043. https://doi.org/10.1016/j.specom.2024.103043

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

Ramírez, A., & Almira Vázquez, A. (2016). Variación del fonema /s/ en contextos de juntura de palabra en el español caleño: una mirada sociolingüística. *Lenguaje*, *44*(1), 11-33. https://doi.org/10.25100/lenguaje.v44i1.4628

Ruiz Vásquez, N. F. (2020). The Dialects of Colombian Spanish: A New Classification. *Lenguaje (Cali, Colombia)*, *48*(2), 160-195. https://doi.org/10.25100/lenguaje.v48i2.8719

Sanders, R. E. (2005). Validating ‘observations’ in discourse studies: a methodological reason for attention to cognition. In H. t. Molder & J. Potter (Eds.), *Conversation and Cognition* (pp. 57-78). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511489990.003

Schegloff, E. A. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. *The American journal of sociology*, *97*(5), 1295-1345. https://doi.org/10.1086/229903

Schegloff, E. A. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in society*, *29*(1), 1-63. https://doi.org/10.1017/S0047404500001019

Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge University Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language (Baltimore)*, *53*(2), 361. https://doi.org/10.2307/413107

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of phonetics*, *39*(1), 96-109. https://doi.org/10.1016/j.wocn.2010.11.006

Searle, J. R. (1969). *Speech acts : an essay in the philosophy of language*. Cambridge U.P.

Senado de la República. (1994). *Servicios públicos domiciliarios (LEY 142 DE 1994)*. Bogotá: Imprenta Nacional

Seyfeddinipur, M., Kita, S., & Indefrey, P. (2008). How speakers interrupt themselves in managing problems in speaking: Evidence from self-repairs. *Cognition*, *108*(3), 837-842. https://doi.org/10.1016/j.cognition.2008.05.004

Sidnell, J., & Barnes, R. (2013). Alternative, subsequent descriptions. In G. Raymond, J. Sidnell, & M. Hayashi (Eds.), *Conversational Repair and Human Understanding* (pp. 322-342). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511757464.011

Steffman, J., Cole, J., & Shattuck-Hufnagel, S. (2024). Intonational categories and continua in American English rising nuclear tunes. *Journal of phonetics*, *104*, 101310. https://doi.org/10.1016/j.wocn.2024.101310

Tuomainen, O., Taschenberger, L., Rosen, S., & Hazan, V. (2022). Speech modifications in interactive speech: Effects of age, sex and noise type. *Philosophical transactions. Biological sciences*, *377*(1841), 20200398-20200398. https://doi.org/10.1098/rstb.2020.0398

Tydgat, I., Diependaele, K., Hartsuiker, R. J., & Pickering, M. J. (2012). How lingering representations of abandoned context words affect speech production. *Acta psychologica*, *140*(3), 218-229. https://doi.org/10.1016/j.actpsy.2012.02.004

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of Native-and Foreign-accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles. *Language and speech*, *53*(4), 510-540. https://doi.org/10.1177/0023830910372495

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, *71*, 147-161. https://doi.org/10.1016/j.wocn.2018.07.008

Vera Diettes, K. J., & Plug, L. (2023). Temporal organization in lexical self-repair in Colombian Spanish unscripted speech. 20th International Congress of Phonetic Sciences, Prague, Czech Republic.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software 4*. https://doi.org/10.21105/joss.01686

Winter, B., & Bürkner, P. C. (2021). Poisson regression for linguists: A tutorial introduction to modelling count data with brms. *Language and linguistics compass*, *15*(11), n/a. https://doi.org/10.1111/lnc3.12439

Winter, B., Fischer, M. H., Scheepers, C., & Myachykov, A. (2023). More is Better: English Language Statistics are Biased Toward Addition. *Cognitive Science*, *47*(4), e13254. https://doi.org/https://doi.org/10.1111/cogs.13254

Appendix A

Language Background and Social Characterization Questionnaire

A.1 Social Characterization Questionnaire (Original)

Cuestionario de caracterización social

A través de este cuestionario se recopilará información para la creación de un perfil social y lingüístico que acompañará los datos obtenidos durante este estudio. El tiempo aproximado para completar el cuestionario es de 10 minutos. Una vez completado el cuestionario, toda su información permanecerá en anonimato. Si tiene alguna inquietud, recuerde que puede contactarnos en los correos que se encuentran en el consentimiento informado, del cual usted recibirá una copia. Le agradecemos de antemano su disposición para completar la información de forma completa y honesta de acuerdo con su conocimiento.

* Obligatoria

1. Escriba el código de identificación que se registró al inicio (e. g. SP_001) *

2. Edad *

3. Género *

4. Lugar de nacimiento *

5. Lugar de residencia (e. g. Bogotá) *

6. Barrio *

7. Estrato social *

8. Por favor, liste los lugares en los que ha residido por más de 6 meses y el tiempo de permanencia (e. g. Bucaramanga, 2 años? *

9. Máximo nivel de escolaridad alcanzado. (e. g. Primaria, bachillerato, estudios universitarios, profesional, posgrado) *

10. Si tiene un título universitario, ¿en qué lugar lo obtuvo? Si tiene más de un título, por favor, indíquenos cuál y el lugar (e. g. Pregrado, especialización, maestría, doctorado, y el lugar). *

11. Ocupación. Si usted es estudiante de tiempo completo, escriba "Estudiante" *

12. ¿Qué idiomas habla? ¿Cómo y cuando los aprendió? *

    Por ejemplo:

    Wayuunaiki. Lo aprendí de nacimiento hablando con mis padres.

    Inglés. Lo aprendí desde los 12 años en el colegio.

13. ¿Qué acento de español colombiano diría usted que habla? *

14. ¿En dónde nacieron sus padres/cuidadores? *

15. ¿En dónde crecieron sus padres/cuidadores? Por favor, liste los lugares en los que ellos vivieron por más de un año. *

16. Ocupación de sus padres/cuidadores. *

17. ¿Qué acento de español colombiano diría usted que sus padres/cuidadores hablan? *

18. ¿Tiene usted algún impedimento auditivo, desorden o patología del habla, o dificultades de aprendizaje (e. g. dislexia)? De ser así, por favor, indíquenos cuáles. *

19. ¿Le gustaría ser invitado a participar en futuros estudios? Si así es, por favor, indíquenos su correo electrónico.

¡Muchas gracias por participar! Al finalizar y enviar, puede cerrar esta ventana.

A.2 Social Characterization Questionnaire (English translation)

This questionnaire collects information to create a social and linguistic profile that will accompany the data obtained during this study. The estimated time to complete the questionnaire is 10 minutes. Once completed, all your information will remain anonymous. If you have any concerns, remember that you can contact us via the emails provided in the informed consent form, of which you will receive a copy. We appreciate your willingness to complete the questionnaire fully and honestly according to your knowledge.

***Required**

Enter the identification code registered at the beginning (e.g., SP_001)*

- ☐____

Age*

- ☐____

Gender*

- ☐____

Place of birth*

- ☐____

Place of residence (e.g., Bogotá)*

- ☐____

Neighbourhood*

- ☐____

Socioeconomic status*

- ☐____

Please list the places where you have lived for more than six months and the duration of your stay (e.g., Bucaramanga, 2 years)*

☐______________________________________

☐______________________________________

☐______________________________________

Highest level of education attained (e.g., Primary school, High school, University studies, Professional degree, Postgraduate)*

☐______________________________________

If you have a university degree, where did you obtain it? If you have more than one degree, please specify which ones and where (e.g., Undergraduate, Specialization, Master's, Doctorate, and location)*

☐______________________________________

☐______________________________________

☐______________________________________

Occupation. If you are a full-time student, write "Student"*

☐______________________________________

What languages do you speak? How and when did you learn them?*

For example:

Wayuunaiki. I learned it from birth by speaking with my parents.

English. I learned it at school when I was 12 years old.

☐______________________________________

☐______________________________________

☐______________________________________

Which Colombian Spanish accent would you say you speak?*

☐______________________________________

Where were your parents/caregivers born?*

☐____

Where did your parents/caregivers grow up? Please list the places where they lived for more than a year.*

☐____

☐____

☐____

Occupation of your parents/caregivers*

☐____

Which Colombian Spanish accent would you say your parents/caregivers speak?*

☐____

Do you have any hearing impairments, speech disorders or conditions, or learning difficulties (e.g., dyslexia)? If so, please specify. *

☐____

Would you like to be invited to participate in future studies? If so, please provide your email address. *

☐____

Thank you very much for participating!

Upon completion and submission, you may close this window.

Appendix B

Ethical Document

The following pages contain the English and Spanish versions of the Participant Information Sheet and Consent Form used in this study. The layout and formatting have been preserved to reflect the original documents.

B.1 Information Sheet (Original)

HOJA DE INFORMACIÓN
Proyecto de investigación: Reparación en el habla conversacional del español colombiano

Usted ha sido invitado a participar en un proyecto de investigación. Antes de decidir si desea participar, es importante que entienda por qué esta investigación se está realizando y sus implicaciones. Por favor, lea cuidadosamente la siguiente información y no dude en preguntar si hay algo que no está claro o si desea más información. Le agradecemos por tomarse el tiempo de leer este documento.

¿Cuál es el propósito de este Proyecto?
Esta investigación tiene como objetivo reunir información sobre cómo los hablantes de los distintos dialectos del español en Colombia formulación y aplican estrategias para corregir distintas instancias de su habla en conversaciones espontáneas mientras al mismo tiempo se recolecta un nuevo corpus del español espontaneo hablado en Colombia.

¿Por qué he sido escogido?
Estamos invitando a participar a hablantes nativos del español colombiano de distintas regiones. Depende de usted decidir si desea o no participar. La negativa a participar no afectará sus derechos de ninguna forma. Si decide participar, se le dará esta hoja de información para que usted la conserve y se le pedirá que firme un formulario de consentimiento. Si decide participar, igualmente usted es libre de retirarse en cualquier momento y no tendrá que dar ninguna justificación.

¿Qué pasará si decido participar?
Primero se le pedirá que complete un cuestionario de caracterización social y, posteriormente, que sostenga una conversación con su compañero(a), en la cual realizaran dos tareas de interacción. Esto no debe tardar más de una hora. Con este fin se le proporcionará algunos temas de conversación y un par de imágenes y se le pedirá que los discuta con su compañero(a) sobre estos.

¿Qué pasa cuando la investigación finaliza?
Si el estudio termina antes de que se haya completado, usted será notificado del porqué. Una vez el proceso de recolección de datos haya finalizado, se empleará un periodo de tiempo en el análisis e interpretación de los datos. La investigación será presentada formalmente a la comunidad académica y otros profesionales. Es importante mencionar que usted podrá ponerse en contacto con nosotros en cualquier momento y una vez finalizada la recolección de datos.

¿Mi participación en este proyecto se mantendrá confidencial?
Toda la información que se recoja sobre usted durante el curso de la investigación se mantendrá estrictamente confidencial. Cualquier información sobre usted que se difunda estará totalmente en el anonimato para que usted no pueda ser identificado a través de ella.

¿Qué pasará con los resultados del proyecto de investigación?
Los resultados aparecerán en la tesis doctoral del investigador principal y harán parte de un repositorio sobre que albergará información sobre los dialectos del español de Colombia. Igualmente, podrán ser compartidos con la

comunidad académica y profesionales del área a través de artículos publicados en revistas académicas o presentaciones en conferencias.

¿Quién está organizando y financiando la investigación?
Esta investigación es llevada a cabo por Kelly Johanna Vera Diettes, Profesora del Departamento de Lingüística de la Universidad Nacional de Colombia y quien hace parte del grupo de investigadores de Posgrado (*Postgraduate Researcher)* del Departamento de Lingüística y Fonética de la Universidad de Leeds (Reino Unido), supervisada por el Dr. Leendert Plug y la Dra. Gisela Tomé Lourido. Este estudio ha recibido aval ético por parte de la Universidad de Leeds y de la Universidad Nacional de Colombia para la investigación con seres humanos

Contacto para mayor información:
Kelly Johanna Vera Diettes
Estudiante de posgrado en investigación, Departamento de Lingüística y Fonética, Escuela de Lenguajes, Culturas y Sociedades, Edificio Michael Sadler, Universidad de Leeds, Leeds, LS2 9JT, UK. / Departamento de Lingüística, Facultad de Ciencias Humanas, Edificio Antonio Nariño (214), of. 231, Bogotá, Colombia. Tel.: (+571) 3165000, ext. 16659
Correo electrónico: ml09kjvd@leeds.ac.uk; kjverad@unal.edu.co

B.2 Information sheet (English Translation)

INFORMATION SHEET
“Research project: Repair in conversational speech from Colombian Spanish”

You have been invited to participate in a research project. Before you can decide if you wish to participate or not, it is important for you to understand why this research is being completed and its implications. Please read the following information carefully and do not hesitate to as if there is anything that is not clear or if there is anything you need more information about.

What is the purpose of this research?
This research aims at investigating how speakers from different Spanish dialects in Colombia formulate and apply strategies to correct different instances of its speech in spontaneous conversations while at the same a new corpus of the spontaneous Spanish spoken in Colombia is collected.

Why have I been chosen?
We are inviting to take part native Spanish speakers from Colombia who come from different regions of the country. It is up to you if you wish to take part or not. Unwillingness to take part will not affect your rights in any way. If you wish to take part, you will be given this Information Sheet for you to keep and you will be asked to complete a Consent Form. Si decide participar, igualmente usted es libre de retirarse en cualquier momento y no tendrá que dar ninguna justificación.

What will happen if I take part?
Firstly, you will be asked to complete a social characterisation questionnaire. Then, we will ask to engage in a conversation with your friend in which you will complete two interactive tasks. This will not take longer than an hour. To that end, you will be given a few conversation topics together with a set of images. You will be asked to discuss with your friend about these pictures.

What happens when the research study stops?
If the study ends before it is completed, you will, of course, be told why. Once the data collection is completed, there will be some time spent in the analysis and interpretation. The research will be formally presented to the academic community and other relevant professionals. You will be able to contact us after the data collection is finished.

Will our taking part in this project be kept confidential?
All information which is collected about you during the course of the research will be kept strictly confidential. Any information about you which is disseminated will be fully anonymised so that you cannot be recognised from it.

What will happen to the results of the research project?
The findings will appear in the lead researcher’s doctoral dissertation and the data will be part of a repository that will keep information about the dialects of the Spanish spoken in Colombia. Additionally, results may later be shared with

the academic and relevant professional communities through articles in academic journals, or presentations at conferences.

Who is organising and funding the research?
This research is being undertaken by Kelly Johanna Vera Diettes, Professor based at the Department of Linguistics at Universidad Nacional de Colombia, who is also a Postgraduate Researcher at the Department of Linguistics and Phonetics at Leeds University, supervised by Dr Leendert Plug y Dr Gisela Tomé Lourido. This study has received ethics clearance through the University of Leeds and Universidad Nacional de Colombia ethical approval processes for research involving human participants.

Contact for further information:
Kelly Johanna Vera Diettes
Department of Linguistics and Phonetics, School of Languages, Cultures and Societies, Michael Sadler Building, University of Leeds, Leeds, LS2 9JT / Departamento de Lingüística, Facultad de Ciencias Humanas, Edificio Antonio Nariño (214), of. 231, Bogotá, Colombia. Tel.: (+571) 3165000, ext. 16659
Email: ml109kjvd@leeds.ac.uk; kjverad@unal.edu.co

B.3 Consent Form (Original)

FORMULARIO DE CONSENTIMIENTO PARA PARTICIPAR EN EL PROYECTO

«Reparación en el habla conversacional del español colombiano»

	Añada sus iniciales al lado del enunciado si está de acuerdo
Confirmo que he leído y entendido la hoja de información de fecha _____________ explicando el proyecto de investigación arriba descrito y he tenido la oportunidad de hacer preguntas sobre el proyecto.	
Entiendo que mi participación es voluntaria y que soy libre de retirarme en cualquier momento sin dar ninguna razón y sin que haya ninguna consecuencia negativa. Si no deseo responder a alguna o varias preguntas, estoy en libertad de rechazarla(s). NOTA: En caso de decidir retirarse, una vez cumplido esto, todos los datos electrónicos se eliminarán de forma permanente y las copias impresas serán destruidas.	
Doy permiso para que los miembros del equipo de investigación tengan acceso a mis respuestas anónimas. Entiendo que mi nombre no estará vinculado con los materiales de investigación, y no voy a ser identificado o identificable en el informe o informes que resulten de la investigación. Entiendo que mis respuestas serán estrictamente confidenciales.	
Estoy de acuerdo con que los datos recolectados de mí sean utilizados en futuras investigaciones pertinentes de forma anónima.	
Estoy de acuerdo en participar en el proyecto de investigación anterior e informaré al investigador principal, si mis datos de contacto cambian.	

Nombre del participante	
Firma del participante	

Fecha	
Nombre del investigador principal	Kelly Johanna Vera Diettes
Nombre de quien llevó a cabo la recolección de datos	Kelly Johanna Vera Diettes
Firma*	
Fecha*	

*Para ser firmado y fechado en presencia del participante

Una vez este consentimiento haya sido firmado por todas las partes, los participantes deben recibir una copia del documento firmado y fechado, de la hoja de información y de cualquier otra información escrita entregada a los participantes. El investigador principal deberá mantener una copia del consentimiento fechado y firmado, junto con los documentos principales de esta investigación en una ubicación segura.

Con el fin de enviarle una copia del presente documento y otra información relevante, por favor, indíquenos su correo electrónico a continuación:

B.4 Consent Form (English Translation)

CONSENT TO TAKE PART IN
"Repair in conversation speech from Colombian Spanish"

	Add your initials next to the statement if you agree
I confirm that I have read and understand the information sheet dated ______________ explaining the above research project and I have had the opportunity to ask questions about the project.	
I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without there being any negative consequences. In addition, should I not wish to answer any particular question or questions, I am free to decline. NB. After withdrawal from the study, all electronic data will be permanently deleted and hard copies destroyed.	
I give permission for members of the research team to have access to my anonymised responses. I understand that my name will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research. I understand that my responses will be kept strictly confidential.	
I agree for the data collected from me to be used in relevant future research in an anonymised form.	
I agree to take part in the above research project and will inform the lead researcher should my contact details change.	

Name of participant	
Participant’s signature	
Date	
Name of lead researcher	Kelly Johanna Vera Diettes
Signature*	Kelly Johanna Vera Diettes
Date*	
Name of participant	

*To be signed and dated in the presence of the participant.

Once this consent form has been signed by all parties, the participant should receive a copy of the signed and dated participant consent form, the information sheet and any other written information provided to the participants. A copy of the signed and dated consent form should be kept by the researcher with the project’s main documents which must be kept in a secure location.

With the aim of sending you a copy of the present document and other relevant information, please, write your email below:

Appendix C

Consensual Response Task: Topics and Questions

C.1 Spanish Version

Vacunación contra el COVID-19

Hasta el momento los procesos de vacunación contra el COVID-19 no han sido obligatorios. Sin embargo, países como Francia, EE. UU. y Canadá han iniciado fuertes campañas que incluyen medidas altamente restrictivas. Por ejemplo, exigir el certificado de vacunación para ingresar a teatros o al transporte público. Colombia adoptó una serie de restricciones similares, ¿creen ustedes que estas son necesarias y han ayudado a combatir la pandemia en el país a través de la vacunación?

Seguridad en las ciudades de Colombia

Alcaldes y alcaldesas de las ciudades de Colombia han implementado distintos tipos de campañas y medidas de seguridad durante los últimos años con el fin de mejorar los niveles de seguridad en las calles. Ante esto, ¿creen ustedes que las iniciativas implementadas han funcionado y las ciudades colombianas son actualmente más seguras que hace cinco años?

iOS y Android

Actualmente tenemos como opción en el mercado teléfonos inteligentes con sistemas operativos que incluyen, en sus últimas versiones, ya sea iOS 15 o Android 12. De una parte, iOS, entre otros, ofrece una integración completa entre software y hardware. Por otro lado, al ser de código abierto Android brinda un gran universo de aplicaciones. ¿Para el 2022, cuál creen ustedes que es la mejor opción para los usuarios en cuanto a, por ejemplo, actualizaciones, mantenimiento, variedad, diseño, seguridad, costo, etc., entre iOS o Android?

Cadena perpetua en Colombia

Recientemente la Corte Constitucional declaró inconstitucional la reforma que reglamentaba la cadena perpetua para violadores de niños. Esto ocurrió a pesar de que en un comienzo la iniciativa había sido bien recibida por algunos sectores de la sociedad debido al constante número de casos denunciados. En este sentido, ¿consideran ustedes que la Corte Constitucional tomó una buena decisión? o, por el contrario, ¿creen la reforma en mención debe ser nuevamente estudiada y la cadena perpetua debería implementarse para este y otros tipos de delitos, inclusive?

Desempeño de la selección Colombia de fútbol y su dirección técnica

A pesar de los esfuerzos, la Selección Colombia no ha clasificado al Mundial del Catar que se realizará a finales de este año. Se ha debatido acerca de las razones de este infortunio; algunos piensan que es responsabilidad del equipo técnico, otros culpan a la Federación Colombia de Fútbol y, además, se ha hablado de la responsabilidad que tienen los jugares mismos sobre este fracaso. ¿De acuerdo con su opinión, cuál o cuáles creen ustedes que son las razones principales que nos llevaron a perdernos el Mundial de Catar?

Presidencia 2022-2026

Teniendo en cuenta que la reciente elección de Gustavo Petro Urrego como presidente de la República, ¿Cuáles creen ustedes deberían ser las prioridades de este gobierno (i. e. Educación, salud, defensa, etc.) teniendo en cuenta el manejo que le dio el anterior presidente a la agenda nacional?

Maltrato animal

Colombia se encuentra a la vanguardia en Latinoamérica en cuanto a políticas que combaten el maltrato animal. Desde hace varios años, en Colombia está prohibido, por ejemplo, tener cautivos animales para incorporarlos en actos circenses. Asimismo, recientemente en el país se aprobó una ley que prohíbe pruebas en animales para la producción de cosméticos e, igualmente, impide la importación a Colombia de productos para hayan sido probados en animales. Teniendo en cuenta que somos uno de los países más biodiversos y como gran variedad de fauna en el planeta, ¿Creen ustedes que en el país se están protegiendo realmente los derechos de los animales? O, por el contrario, ¿debemos avanzar más en este aspecto? De ser así, ¿Qué otras medidas se podrían incorporar?

C.1 English Translation

COVID-19 Vaccination

Up to this point, COVID-19 vaccination campaigns have not been mandatory. However, countries like France, the United States, and Canada have implemented strong campaigns involving highly restrictive measures. For example, requiring proof of vaccination to enter theaters or use public transportation. Colombia adopted a series of similar restrictions. Do you believe these measures are necessary and have helped combat the pandemic through vaccination in the country?

Security in Colombian Cities

Mayors across Colombia have implemented various security measures and public safety campaigns in recent years to improve conditions in urban areas. In your opinion, have these initiatives been effective? Are Colombian cities safer today than they were five years ago?

iOS vs Android

Today’s smartphone market offers devices running the latest versions of either iOS 15 or Android 12. On one hand, iOS offers complete integration between software and hardware. On the other hand, Android’s open-source nature allows access to a vast universe of applications. For 2022, which do you think is the better option for users in terms of updates, maintenance, variety, design, security, cost, etc. — iOS or Android?

Life Imprisonment in Colombia

Recently, the Constitutional Court struck down the reform that would have legalized life imprisonment for child molesters. This decision came despite the fact that the initiative was initially well received by certain sectors of society, due to the persistently high number of reported cases. In this context, do you believe the Constitutional Court made the right decision? Or, do you think the reform should be reconsidered and life imprisonment implemented for this and possibly other types of crimes?

Performance of the Colombian National Football Team and Its Coaching Staff

Despite various efforts, the Colombian national football team failed to qualify for the 2022 World Cup in Qatar. Several explanations have been debated: some blame the coaching staff, others point to the Colombian Football Federation, and some have attributed the outcome to the players themselves. In your opinion, what do you think were the main reasons Colombia did not qualify for the World Cup?

2022–2026 Presidency

Considering the recent election of Gustavo Petro Urrego as President of the Republic, what do you believe should be the priorities of this new government (e.g., education, health, defense, etc.), especially in light of how the previous president managed the national agenda?

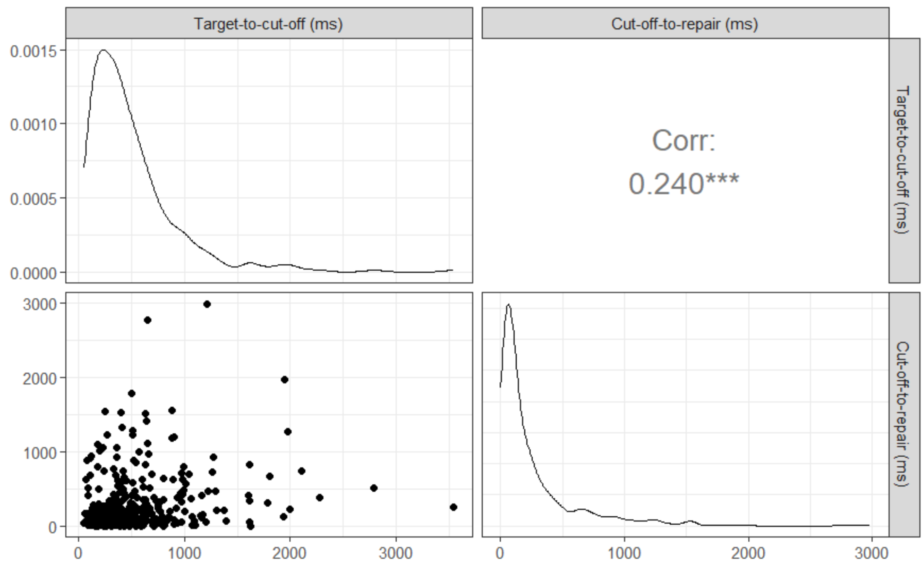
Animal Cruelty

Colombia is among the leading countries in Latin America in terms of legislation against animal cruelty. For example, animals have long been banned from use in circuses, and more recently, a law was passed prohibiting animal testing for cosmetic production, as well as the importation of any cosmetic products tested on animals. Given that Colombia is one of the world's most biodiverse countries, do you believe animal rights are being adequately protected? Or should the country go further? If so, what additional measures could be implemented?

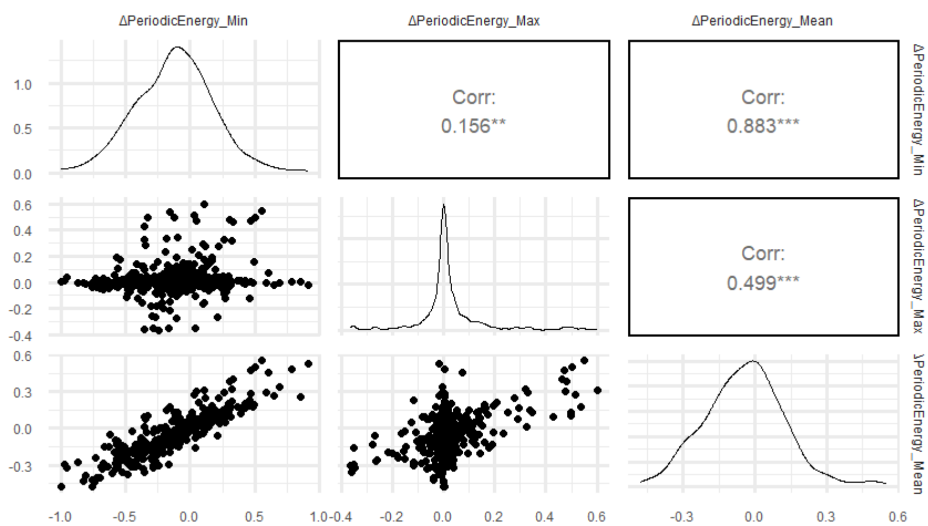
Appendix D

Correlation Tests

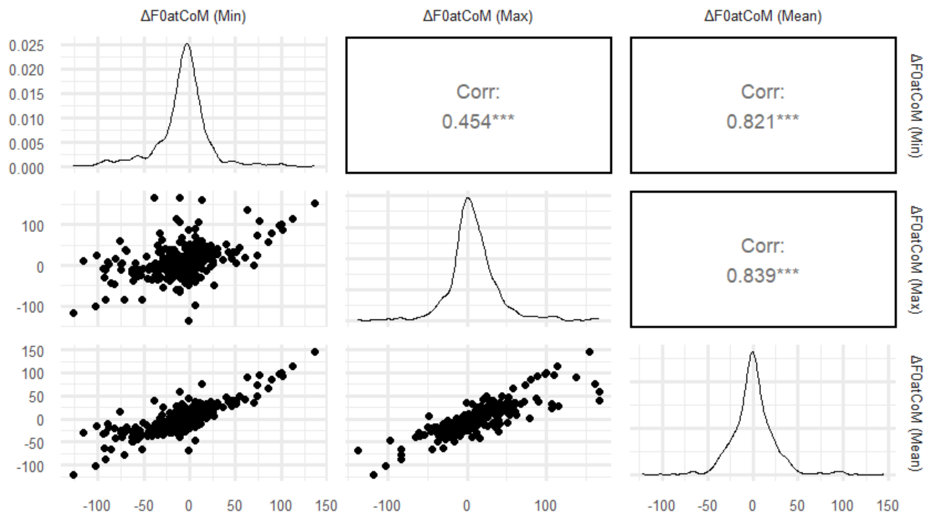
D.1 Scatterplot Matrix for Target-to-cut-off and Cut-off-to-repair



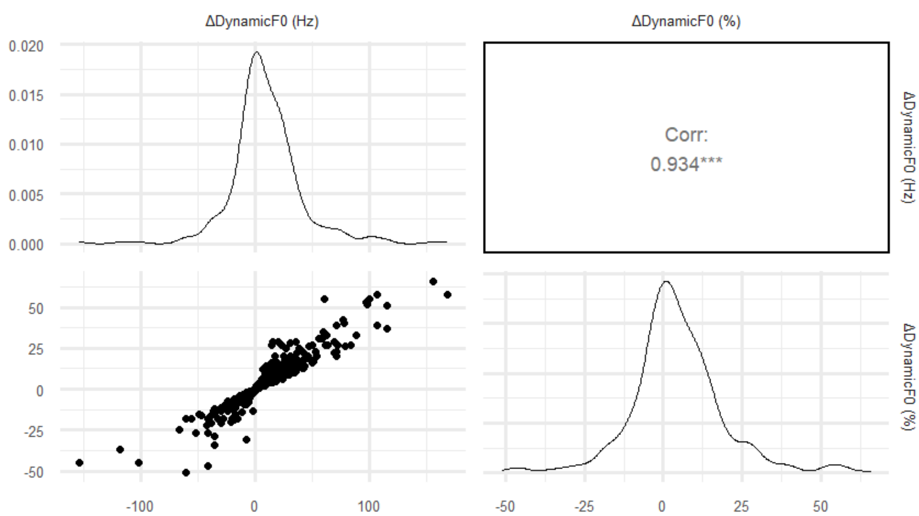
D.2 Scatterplot Matrix for Periodic Energy



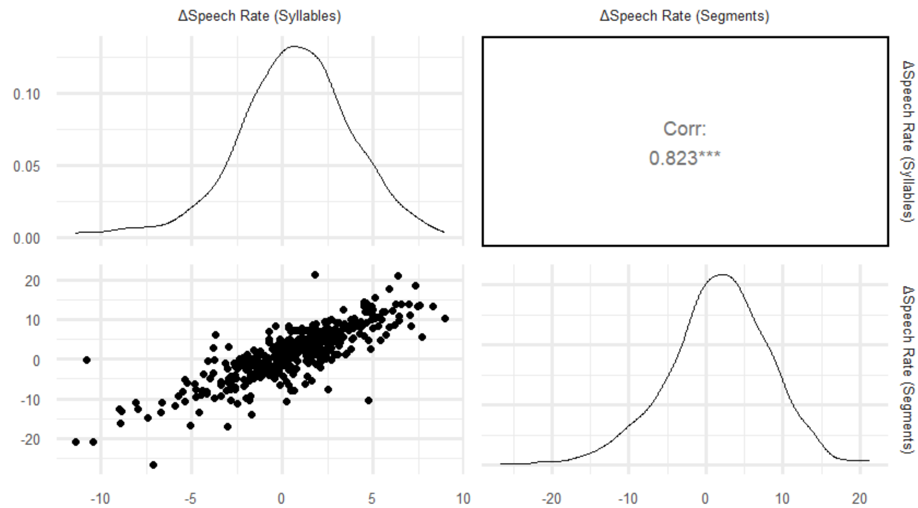
D.3 Scatterplot Matrix for F0 at CoM



D.4 Scatterplot Matrix for Dynamic F0



D.5 Scatterplot Matrix for Speech Rate



Appendix E

Bayesian Modelling Results

E.1 Summary of the Model Examining Target-to-cut-off (Log-transformed)

```
Family: lognormal
  Links: mu = identity; sigma = identity
Formula: TargToCutOff ~ 1 + Completeness + Ed_Term + Err_Type + Err_Type:Ed_Term + (1
+ Completeness + Ed_Term + Err_Type + Err_Type:Ed_Term | Speaker)
   Data: timing_df (Number of observations: 404)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                                              Estimate Est.Error l-95% CI u-95% CI
sd(Intercept)                                     0.09      0.07     0.00     0.22
sd(CompletenessI)                                 0.06      0.08     0.00     0.27
sd(Ed_TermYES)                                    0.06      0.09     0.00     0.30
sd(Err_TypeF)                                     0.04      0.06     0.00     0.22
sd(Err_TypeL)                                     0.02      0.04     0.00     0.13
sd(Ed_TermYES:Err_TypeF)                          0.02      0.04     0.00     0.13
sd(Ed_TermYES:Err_TypeL)                          0.09      0.22     0.00     0.85
cor(Intercept,CompletenessI)                      0.04      0.31    -0.57     0.64
cor(Intercept,Ed_TermYES)                         0.01      0.31    -0.58     0.60
cor(CompletenessI,Ed_TermYES)                     0.02      0.31    -0.59     0.62
cor(Intercept,Err_TypeF)                          0.03      0.32    -0.59     0.62
cor(CompletenessI,Err_TypeF)                      0.02      0.31    -0.58     0.61
cor(Ed_TermYES,Err_TypeF)                         0.01      0.32    -0.60     0.60
cor(Intercept,Err_TypeL)                         -0.01      0.32    -0.61     0.60
cor(CompletenessI,Err_TypeL)                     -0.00      0.32    -0.61     0.60
cor(Ed_TermYES,Err_TypeL)                         0.00      0.32    -0.60     0.60
cor(Err_TypeF,Err_TypeL)                          0.00      0.32    -0.60     0.61
cor(Intercept,Ed_TermYES:Err_TypeF)              -0.01      0.32    -0.61     0.61
cor(CompletenessI,Ed_TermYES:Err_TypeF)           0.00      0.32    -0.60     0.61
cor(Ed_TermYES,Ed_TermYES:Err_TypeF)              0.00      0.32    -0.60     0.61
cor(Err_TypeF,Ed_TermYES:Err_TypeF)              -0.00      0.32    -0.61     0.60
cor(Err_TypeL,Ed_TermYES:Err_TypeF)               0.00      0.32    -0.60     0.60
cor(Intercept,Ed_TermYES:Err_TypeL)               0.00      0.31    -0.60     0.60
cor(CompletenessI,Ed_TermYES:Err_TypeL)          -0.00      0.32    -0.59     0.60
cor(Ed_TermYES,Ed_TermYES:Err_TypeL)              0.00      0.32    -0.61     0.60
cor(Err_TypeF,Ed_TermYES:Err_TypeL)               0.00      0.31    -0.59     0.60
cor(Err_TypeL,Ed_TermYES:Err_TypeL)              -0.00      0.32    -0.61     0.60
cor(Ed_TermYES:Err_TypeF,Ed_TermYES:Err_TypeL)   -0.00      0.32    -0.61     0.60
                                              Rhat Bulk_ESS Tail_ESS
sd(Intercept)                                 1.00     1823     6091
sd(CompletenessI)                             1.00     3181     4760
sd(Ed_TermYES)                                1.00     4238     5515
sd(Err_TypeF)                                 1.00     4720     5583
sd(Err_TypeL)                                 1.00    17251    10051
sd(Ed_TermYES:Err_TypeF)                      1.00    18195     9312
sd(Ed_TermYES:Err_TypeL)                      1.00     6112     3817
cor(Intercept,CompletenessI)                  1.00    19505    10945
cor(Intercept,Ed_TermYES)                     1.00    25368    11665
cor(CompletenessI,Ed_TermYES)                 1.00    19678    11995
cor(Intercept,Err_TypeF)                      1.00    25498    11446
cor(CompletenessI,Err_TypeF)                  1.00    19240    11586
cor(Ed_TermYES,Err_TypeF)                     1.00    17397    11716
cor(Intercept,Err_TypeL)                      1.00    29287    10906
cor(CompletenessI,Err_TypeL)                  1.00    22158    11268
cor(Ed_TermYES,Err_TypeL)                     1.00    17851    11808
cor(Err_TypeF,Err_TypeL)                      1.00    13967    11449
cor(Intercept,Ed_TermYES:Err_TypeF)           1.00    27403    10527
cor(CompletenessI,Ed_TermYES:Err_TypeF)       1.00    22953    11846
```

```
cor(Ed_TermYES,Ed_TermYES:Err_TypeF)           1.00    17107    11334
cor(Err_TypeF,Ed_TermYES:Err_TypeF)            1.00    15106    12700
cor(Err_TypeL,Ed_TermYES:Err_TypeF)            1.00    12164    11640
cor(Intercept,Ed_TermYES:Err_TypeL)            1.00    29387    10049
cor(CompletenessI,Ed_TermYES:Err_TypeL)        1.00    21877    11301
cor(Ed_TermYES,Ed_TermYES:Err_TypeL)           1.00    17597    11973
cor(Err_TypeF,Ed_TermYES:Err_TypeL)            1.00    14986    12223
cor(Err_TypeL,Ed_TermYES:Err_TypeL)            1.00    11452    11545
cor(Ed_TermYES:Err_TypeF,Ed_TermYES:Err_TypeL) 1.00    10217    11704

Regression Coefficients:
                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept                6.19      0.07     6.06     6.33 1.00    17858    12608
CompletenessI           -0.67      0.07    -0.81    -0.54 1.00    20678    11224
Ed_TermYES               0.50      0.11     0.27     0.72 1.00    13478    11722
Err_TypeF               -0.11      0.08    -0.27     0.05 1.00    16092    12948
Err_TypeL               -0.15      0.10    -0.34     0.05 1.00    19553    12584
Ed_TermYES:Err_TypeF    -0.19      0.15    -0.48     0.10 1.00    13210    12686
Ed_TermYES:Err_TypeL    -0.03      0.24    -0.50     0.44 1.00    18976    12376

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.64      0.02     0.59     0.69 1.00    16753    12153

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.2. Summary of the Model Examining Cut-off-to-repair (Log-transformed)

```
Family: lognormal
  Links: mu = identity; sigma = identity
Formula: CutOffToRep ~ 1 + Completeness + Ed_Term + Err_Type + (1 + Completeness +
Ed_Term + Err_Type | Speaker)
   Data: timing_df (Number of observations: 404)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.10      0.13     0.00     0.42 1.00     1997
sd(CompletenessI)                 0.03      0.05     0.00     0.18 1.00    15858
sd(Ed_TermYES)                    0.02      0.03     0.00     0.10 1.00    26226
sd(Err_TypeF)                     0.04      0.10     0.00     0.38 1.00     6555
sd(Err_TypeL)                     0.02      0.04     0.00     0.13 1.00    26060
cor(Intercept,CompletenessI)      0.00      0.35    -0.66     0.66 1.00    30152
cor(Intercept,Ed_TermYES)        -0.01      0.35    -0.67     0.66 1.00    29514
cor(CompletenessI,Ed_TermYES)    -0.00      0.35    -0.66     0.67 1.00    19979
cor(Intercept,Err_TypeF)          0.01      0.35    -0.66     0.67 1.00    24831
cor(CompletenessI,Err_TypeF)      0.00      0.35    -0.67     0.67 1.00    18521
cor(Ed_TermYES,Err_TypeF)        -0.00      0.36    -0.67     0.67 1.00    15421
cor(Intercept,Err_TypeL)         -0.00      0.35    -0.67     0.66 1.00    29289
cor(CompletenessI,Err_TypeL)     -0.01      0.36    -0.67     0.66 1.00    20971
cor(Ed_TermYES,Err_TypeL)         0.00      0.35    -0.66     0.67 1.00    15167
cor(Err_TypeF,Err_TypeL)          0.00      0.35    -0.66     0.66 1.00    13212
                              Tail_ESS
sd(Intercept)                     3825
sd(CompletenessI)                 8512
sd(Ed_TermYES)                    9848
sd(Err_TypeF)                     3479
sd(Err_TypeL)                     8920
cor(Intercept,CompletenessI)     11079
cor(Intercept,Ed_TermYES)        11273
cor(CompletenessI,Ed_TermYES)    12190
cor(Intercept,Err_TypeF)         10616
cor(CompletenessI,Err_TypeF)     11081
cor(Ed_TermYES,Err_TypeF)        11219
cor(Intercept,Err_TypeL)         10841
```

```
cor(CompletenessI,Err_TypeL)      11356
cor(Ed_TermYES,Err_TypeL)         11837
cor(Err_TypeF,Err_TypeL)          11485

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept         4.40      0.12     4.16     4.65 1.00    16812    11337
CompletenessI    -0.22      0.13    -0.46     0.03 1.00    29127    11815
Ed_TermYES        1.57      0.14     1.29     1.85 1.00    28647    10884
Err_TypeF         0.05      0.14    -0.22     0.32 1.00    20607    12301
Err_TypeL        -0.08      0.18    -0.43     0.27 1.00    26035    13185

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.29      0.05     1.19     1.39 1.00    10026    10505

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.3 Summary of the Model Examining Periodic Energy Minimum

```
Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Delta_Min_PeriodicEnergy ~ 1 + Completeness + Ed_Term + Err_Type + (1 +
Completeness + Ed_Term + Err_Type | Speaker)
   Data: df_PeriodicEnergy (Number of observations: 405)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                    0.03      0.02     0.00     0.07 1.00     7503
sd(CompletenessI)                0.04      0.03     0.00     0.12 1.00     5648
sd(Ed_TermYES)                   0.09      0.05     0.00     0.20 1.00     3640
sd(Err_TypeF)                    0.04      0.03     0.00     0.11 1.00     6249
sd(Err_TypeL)                    0.06      0.05     0.00     0.18 1.00     6156
cor(Intercept,CompletenessI)    -0.03      0.35    -0.69     0.65 1.00    14621
cor(Intercept,Ed_TermYES)       -0.07      0.35    -0.70     0.61 1.00     7723
cor(CompletenessI,Ed_TermYES)   -0.00      0.35    -0.65     0.66 1.00    10116
cor(Intercept,Err_TypeF)        -0.02      0.35    -0.69     0.65 1.00    15130
cor(CompletenessI,Err_TypeF)    -0.02      0.35    -0.68     0.66 1.00    13577
cor(Ed_TermYES,Err_TypeF)       -0.04      0.35    -0.67     0.63 1.00    14183
cor(Intercept,Err_TypeL)        -0.02      0.36    -0.68     0.65 1.00    14612
cor(CompletenessI,Err_TypeL)     0.02      0.35    -0.64     0.67 1.00    13756
cor(Ed_TermYES,Err_TypeL)        0.03      0.34    -0.63     0.67 1.00    13954
cor(Err_TypeF,Err_TypeL)         0.03      0.35    -0.64     0.68 1.00    12151
                             Tail_ESS
sd(Intercept)                    7332
sd(CompletenessI)                6118
sd(Ed_TermYES)                   4645
sd(Err_TypeF)                    7208
sd(Err_TypeL)                    7110
cor(Intercept,CompletenessI)    11779
cor(Intercept,Ed_TermYES)       10427
cor(CompletenessI,Ed_TermYES)   11928
cor(Intercept,Err_TypeF)        11483
cor(CompletenessI,Err_TypeF)    11724
cor(Ed_TermYES,Err_TypeF)       12705
cor(Intercept,Err_TypeL)        11958
cor(CompletenessI,Err_TypeL)    12742
cor(Ed_TermYES,Err_TypeL)       11960
cor(Err_TypeF,Err_TypeL)        12580

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        -0.13      0.03    -0.18    -0.07 1.00    19659    12699
```

```
CompletenessI    -0.14      0.03    -0.20    -0.07 1.00    16623    12182
Ed_TermYES        0.05      0.04    -0.03     0.13 1.00    13040    10752
Err_TypeF         0.07      0.03     0.01     0.14 1.00    15223    12045
Err_TypeL         0.11      0.05     0.02     0.20 1.00    14639    11770

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.29      0.01     0.27     0.31 1.00    13955    11100

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.4 Summary of the Model Examining Periodic Energy Mean

```
Family: gaussian
  Links: mu = identity; sigma = identity
Formula: Delta_Mean_PeriodicEnergy ~ 1 + Completeness + Ed_Term + Err_Type + (1 +
Completeness + Ed_Term + Err_Type | Speaker)
   Data: df_PeriodicEnergy (Number of observations: 405)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.02      0.01     0.00     0.05 1.00     7619
sd(CompletenessI)                 0.04      0.02     0.00     0.09 1.00     4701
sd(Ed_TermYES)                    0.05      0.03     0.00     0.12 1.00     3954
sd(Err_TypeF)                     0.02      0.01     0.00     0.06 1.00     9233
sd(Err_TypeL)                     0.04      0.03     0.00     0.10 1.00     7921
cor(Intercept,CompletenessI)     -0.00      0.35    -0.66     0.65 1.00    12701
cor(Intercept,Ed_TermYES)        -0.07      0.35    -0.70     0.61 1.00    10930
cor(CompletenessI,Ed_TermYES)    -0.03      0.34    -0.67     0.62 1.00    12217
cor(Intercept,Err_TypeF)         -0.03      0.35    -0.69     0.65 1.00    23961
cor(CompletenessI,Err_TypeF)     -0.02      0.35    -0.68     0.66 1.00    19188
cor(Ed_TermYES,Err_TypeF)        -0.03      0.36    -0.69     0.65 1.00    19921
cor(Intercept,Err_TypeL)         -0.00      0.35    -0.67     0.67 1.00    23575
cor(CompletenessI,Err_TypeL)      0.03      0.35    -0.64     0.68 1.00    19747
cor(Ed_TermYES,Err_TypeL)         0.03      0.35    -0.65     0.67 1.00    17836
cor(Err_TypeF,Err_TypeL)          0.03      0.35    -0.65     0.68 1.00    12063
                              Tail_ESS
sd(Intercept)                     9086
sd(CompletenessI)                 6493
sd(Ed_TermYES)                    5342
sd(Err_TypeF)                     8665
sd(Err_TypeL)                     7969
cor(Intercept,CompletenessI)     12495
cor(Intercept,Ed_TermYES)        11191
cor(CompletenessI,Ed_TermYES)    12469
cor(Intercept,Err_TypeF)         12288
cor(CompletenessI,Err_TypeF)     12525
cor(Ed_TermYES,Err_TypeF)        12931
cor(Intercept,Err_TypeL)         12274
cor(CompletenessI,Err_TypeL)     12851
cor(Ed_TermYES,Err_TypeL)        13345
cor(Err_TypeF,Err_TypeL)         12391

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        -0.05      0.02    -0.08    -0.02 1.00    31318    12782
CompletenessI    -0.05      0.02    -0.09    -0.01 1.00    20798    12286
Ed_TermYES        0.02      0.02    -0.02     0.07 1.00    16810    10894
Err_TypeF         0.04      0.02    -0.00     0.08 1.00    25112    13122
Err_TypeL         0.05      0.03    -0.01     0.10 1.00    21094    13713

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.16      0.01     0.15     0.18 1.00    19449    11898
```

```
Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.5 Summary of the Model Examining Periodic Energy Maximum

```
Family: student
  Links: mu = identity; sigma = identity; nu = identity
Formula: Delta_Max_PeriodicEnergy ~ 1 + Completeness + Ed_Term + Err_Type + (1 +
Completeness + Ed_Term + Err_Type | Speaker)
   Data: df_filtered_MaxPeriodicEnergy_45 (Number of observations: 375)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                      0.00      0.00     0.00     0.01 1.00    11065
sd(CompletenessI)                  0.01      0.01     0.00     0.03 1.00     3160
sd(Ed_TermYES)                     0.01      0.01     0.00     0.02 1.00     5205
sd(Err_TypeF)                      0.01      0.00     0.00     0.01 1.00     5635
sd(Err_TypeL)                      0.01      0.01     0.00     0.02 1.00     6348
cor(Intercept,CompletenessI)      -0.02      0.35    -0.68     0.64 1.00    11827
cor(Intercept,Ed_TermYES)         -0.03      0.36    -0.70     0.64 1.00    15523
cor(CompletenessI,Ed_TermYES)      0.06      0.35    -0.62     0.69 1.00    13653
cor(Intercept,Err_TypeF)          -0.06      0.36    -0.71     0.64 1.00    16308
cor(CompletenessI,Err_TypeF)      -0.01      0.35    -0.67     0.66 1.00    16801
cor(Ed_TermYES,Err_TypeF)         -0.10      0.36    -0.75     0.61 1.00    12722
cor(Intercept,Err_TypeL)          -0.05      0.35    -0.70     0.62 1.00    17720
cor(CompletenessI,Err_TypeL)      -0.04      0.36    -0.70     0.65 1.00    18387
cor(Ed_TermYES,Err_TypeL)          0.05      0.35    -0.63     0.70 1.00    14594
cor(Err_TypeF,Err_TypeL)          -0.04      0.35    -0.69     0.62 1.00    13030
                               Tail_ESS
sd(Intercept)                      8574
sd(CompletenessI)                  5354
sd(Ed_TermYES)                     7209
sd(Err_TypeF)                      9667
sd(Err_TypeL)                      6490
cor(Intercept,CompletenessI)      12397
cor(Intercept,Ed_TermYES)         12147
cor(CompletenessI,Ed_TermYES)     12456
cor(Intercept,Err_TypeF)          12268
cor(CompletenessI,Err_TypeF)      12128
cor(Ed_TermYES,Err_TypeF)         12435
cor(Intercept,Err_TypeL)          12410
cor(CompletenessI,Err_TypeL)      13170
cor(Ed_TermYES,Err_TypeL)         12630
cor(Err_TypeF,Err_TypeL)          12610

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept         0.00      0.00    -0.00     0.01 1.00    27358    13102
CompletenessI    -0.00      0.00    -0.01     0.00 1.00    13337     9564
Ed_TermYES        0.00      0.00    -0.01     0.01 1.00    18580    12217
Err_TypeF         0.00      0.00    -0.01     0.01 1.00    19270    11879
Err_TypeL         0.01      0.01    -0.00     0.02 1.00    16459    12520

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     0.02      0.00     0.02     0.03 1.00    12002    10078
nu        1.44      0.21     1.09     1.91 1.00     9968     6346

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.6 Summary of the Model Examining Minimum F0 at CoM

```
Family: student 
  Links: mu = identity; sigma = identity; nu = identity 
Formula: Delta_Min_f0atCoM ~ 1 + Completeness + Ed_Term + Err_Type + (1 + Completeness 
+ Ed_Term + Err_Type | Speaker) 
   Data: df_filtered_MinF0_40 (Number of observations: 398) 
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36) 
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.27      0.53     0.00     1.87 1.00    11351
sd(CompletenessI)                 0.28      0.59     0.00     2.02 1.00    14621
sd(Ed_TermYES)                    0.22      0.40     0.00     1.30 1.00    18923
sd(Err_TypeF)                     0.22      0.40     0.00     1.27 1.00    18315
sd(Err_TypeL)                     0.24      0.49     0.00     1.50 1.00    17320
cor(Intercept,CompletenessI)     -0.00      0.36    -0.67     0.68 1.00    23538
cor(Intercept,Ed_TermYES)         0.00      0.35    -0.66     0.67 1.00    23476
cor(CompletenessI,Ed_TermYES)    -0.00      0.35    -0.67     0.66 1.00    17182
cor(Intercept,Err_TypeF)          0.00      0.35    -0.67     0.67 1.00    21813
cor(CompletenessI,Err_TypeF)     -0.00      0.35    -0.66     0.67 1.00    16555
cor(Ed_TermYES,Err_TypeF)        -0.01      0.35    -0.67     0.66 1.00    14966
cor(Intercept,Err_TypeL)          0.00      0.35    -0.66     0.66 1.00    23460
cor(CompletenessI,Err_TypeL)      0.00      0.35    -0.66     0.67 1.00    16929
cor(Ed_TermYES,Err_TypeL)         0.00      0.35    -0.67     0.67 1.00    15066
cor(Err_TypeF,Err_TypeL)          0.00      0.36    -0.67     0.67 1.00    12947
                              Tail_ESS
sd(Intercept)                     6687
sd(CompletenessI)                 8678
sd(Ed_TermYES)                    7891
sd(Err_TypeF)                     8692
sd(Err_TypeL)                     7826
cor(Intercept,CompletenessI)     11352
cor(Intercept,Ed_TermYES)        11204
cor(CompletenessI,Ed_TermYES)    11495
cor(Intercept,Err_TypeF)         10503
cor(CompletenessI,Err_TypeF)     11586
cor(Ed_TermYES,Err_TypeF)        12153
cor(Intercept,Err_TypeL)         11608
cor(CompletenessI,Err_TypeL)     11925
cor(Ed_TermYES,Err_TypeL)        11681
cor(Err_TypeF,Err_TypeL)         12333

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        -3.97      0.95    -5.90    -2.13 1.00    19496    12131
CompletenessI    -0.21      0.48    -1.15     0.74 1.00    23584    12098
Ed_TermYES        0.02      0.48    -0.93     0.97 1.00    23253    11574
Err_TypeF        -0.05      0.48    -0.98     0.89 1.00    22178    12070
Err_TypeL         0.10      0.49    -0.88     1.07 1.00    24943    11979

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    13.59      1.11    11.53    15.85 1.00    14502    12222
nu        1.91      0.27     1.45     2.51 1.00    14274    10414

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.7 Summary of the Model Examining Mean F0 at CoM

```
Family: student
  Links: mu = identity; sigma = identity; nu = identity
Formula: Delta_Mean_f0atCoM ~ 1 + Completeness + Ed_Term + Err_Type + (1 +
Completeness + Ed_Term + Err_Type | Speaker)
   Data: df_ProsodicMarking (Number of observations: 405)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.28      0.55     0.00     1.89 1.00     9929
sd(CompletenessI)                 0.26      0.57     0.00     1.77 1.00    14522
sd(Ed_TermYES)                    0.25      0.51     0.00     1.56 1.00    16364
sd(Err_TypeF)                     0.22      0.41     0.00     1.37 1.00    19843
sd(Err_TypeL)                     0.26      0.60     0.00     1.69 1.00    17622
cor(Intercept,CompletenessI)      0.00      0.35    -0.66     0.67 1.00    22401
cor(Intercept,Ed_TermYES)        -0.00      0.35    -0.66     0.65 1.00    22559
cor(CompletenessI,Ed_TermYES)    -0.00      0.35    -0.65     0.65 1.00    16251
cor(Intercept,Err_TypeF)          0.00      0.35    -0.67     0.66 1.00    23902
cor(CompletenessI,Err_TypeF)      0.00      0.35    -0.67     0.66 1.00    16457
cor(Ed_TermYES,Err_TypeF)        -0.00      0.35    -0.68     0.66 1.00    15501
cor(Intercept,Err_TypeL)          0.00      0.35    -0.66     0.67 1.00    22124
cor(CompletenessI,Err_TypeL)      0.01      0.35    -0.67     0.67 1.00    18032
cor(Ed_TermYES,Err_TypeL)        -0.00      0.35    -0.66     0.66 1.00    15687
cor(Err_TypeF,Err_TypeL)         -0.00      0.35    -0.66     0.67 1.00    11856
                              Tail_ESS
sd(Intercept)                     5236
sd(CompletenessI)                 7969
sd(Ed_TermYES)                    8914
sd(Err_TypeF)                     8642
sd(Err_TypeL)                     8986
cor(Intercept,CompletenessI)     11783
cor(Intercept,Ed_TermYES)        12247
cor(CompletenessI,Ed_TermYES)    11211
cor(Intercept,Err_TypeF)         11640
cor(CompletenessI,Err_TypeF)     12312
cor(Ed_TermYES,Err_TypeF)        12599
cor(Intercept,Err_TypeL)         10913
cor(CompletenessI,Err_TypeL)     11686
cor(Ed_TermYES,Err_TypeL)        11984
cor(Err_TypeF,Err_TypeL)         11230

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        -0.33      0.98    -2.28     1.59 1.00    22266    12154
CompletenessI     0.23      0.48    -0.70     1.17 1.00    22636    12237
Ed_TermYES       -0.11      0.48    -1.05     0.85 1.00    21744    12099
Err_TypeF        -0.07      0.48    -1.01     0.87 1.00    22467    12319
Err_TypeL         0.07      0.49    -0.89     1.05 1.00    22953    11945

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    14.71      1.09    12.64    16.90 1.00    14071    12327
nu        2.43      0.38     1.79     3.28 1.00    14363    11846

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.8 Summary of the Model Examining Maximum F0 at CoM

```
Family: student
  Links: mu = identity; sigma = identity; nu = identity
```

```
Formula: Delta_Max_f0atCoM ~ 1 + Completeness + Ed_Term + Err_Type + (1 + Completeness
+ Ed_Term + Err_Type | Speaker)
   Data: df_ProsodicMarking (Number of observations: 405)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.27      0.53     0.00     1.90 1.00    20095
sd(CompletenessI)                 0.79      1.86     0.00     7.27 1.00     6032
sd(Ed_TermYES)                    0.26      0.61     0.00     1.65 1.00    28607
sd(Err_TypeF)                     0.22      0.42     0.00     1.32 1.00    30079
sd(Err_TypeL)                     0.27      0.64     0.00     1.79 1.00    30159
cor(Intercept,CompletenessI)      0.00      0.35    -0.66     0.66 1.00    33910
cor(Intercept,Ed_TermYES)        -0.00      0.35    -0.66     0.66 1.00    42989
cor(CompletenessI,Ed_TermYES)     0.00      0.35    -0.66     0.66 1.00    24525
cor(Intercept,Err_TypeF)          0.00      0.36    -0.67     0.68 1.00    44921
cor(CompletenessI,Err_TypeF)     -0.01      0.36    -0.67     0.67 1.00    25106
cor(Ed_TermYES,Err_TypeF)        -0.00      0.35    -0.66     0.67 1.00    17734
cor(Intercept,Err_TypeL)          0.00      0.35    -0.66     0.66 1.00    40360
cor(CompletenessI,Err_TypeL)     -0.00      0.35    -0.67     0.66 1.00    25027
cor(Ed_TermYES,Err_TypeL)         0.00      0.35    -0.67     0.66 1.00    17031
cor(Err_TypeF,Err_TypeL)          0.00      0.36    -0.67     0.66 1.00    12296
                              Tail_ESS
sd(Intercept)                     8870
sd(CompletenessI)                 3590
sd(Ed_TermYES)                    9613
sd(Err_TypeF)                     9569
sd(Err_TypeL)                     8484
cor(Intercept,CompletenessI)     11678
cor(Intercept,Ed_TermYES)        10181
cor(CompletenessI,Ed_TermYES)    11883
cor(Intercept,Err_TypeF)         10815
cor(CompletenessI,Err_TypeF)     11209
cor(Ed_TermYES,Err_TypeF)        12427
cor(Intercept,Err_TypeL)         10302
cor(CompletenessI,Err_TypeL)     10616
cor(Ed_TermYES,Err_TypeL)        11158
cor(Err_TypeF,Err_TypeL)         11554

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept         5.23      1.21     2.84     7.56 1.00    17277    10847
CompletenessI     0.45      0.49    -0.52     1.40 1.00    38856    10982
Ed_TermYES       -0.12      0.50    -1.10     0.85 1.00    41407    10367
Err_TypeF        -0.19      0.49    -1.16     0.77 1.00    38043    11073
Err_TypeL        -0.01      0.49    -0.99     0.94 1.00    41291    10404

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    17.38      1.26    14.99    19.93 1.00    23025    13127
nu        2.14      0.30     1.64     2.79 1.00    24783    12813

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.9 Summary of the Model Examining Relative F0 Rise (%)

```
Family: student
  Links: mu = identity; sigma = identity; nu = identity
Formula: max_DeltaF0_Percentage ~ 1 + Completeness + Ed_Term + Err_Type + (1 +
Completeness + Ed_Term + Err_Type | Speaker)
   Data: df_ProsodicMarking (Number of observations: 405)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000
```

```
Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.18      0.27     0.00     0.95 1.00    18515
sd(CompletenessI)                 0.37      0.82     0.00     3.09 1.00     8587
sd(Ed_TermYES)                    0.26      0.52     0.00     1.85 1.00    16554
sd(Err_TypeF)                     0.21      0.37     0.00     1.27 1.00    19548
sd(Err_TypeL)                     0.24      0.45     0.00     1.45 1.00    20880
cor(Intercept,CompletenessI)      0.00      0.36    -0.66     0.67 1.00    27526
cor(Intercept,Ed_TermYES)        -0.00      0.35    -0.66     0.66 1.00    29810
cor(CompletenessI,Ed_TermYES)    -0.01      0.35    -0.66     0.66 1.00    21586
cor(Intercept,Err_TypeF)          0.00      0.35    -0.67     0.67 1.00    32735
cor(CompletenessI,Err_TypeF)     -0.00      0.35    -0.66     0.66 1.00    21968
cor(Ed_TermYES,Err_TypeF)        -0.00      0.35    -0.67     0.67 1.00    15162
cor(Intercept,Err_TypeL)         -0.00      0.35    -0.66     0.66 1.00    34612
cor(CompletenessI,Err_TypeL)     -0.00      0.35    -0.66     0.66 1.00    21350
cor(Ed_TermYES,Err_TypeL)        -0.00      0.35    -0.67     0.66 1.00    15842
cor(Err_TypeF,Err_TypeL)         -0.00      0.35    -0.67     0.67 1.00    11768
                              Tail_ESS
sd(Intercept)                     9403
sd(CompletenessI)                 4187
sd(Ed_TermYES)                    8655
sd(Err_TypeF)                    10649
sd(Err_TypeL)                     9669
cor(Intercept,CompletenessI)     11057
cor(Intercept,Ed_TermYES)        10910
cor(CompletenessI,Ed_TermYES)    11106
cor(Intercept,Err_TypeF)         10779
cor(CompletenessI,Err_TypeF)     12054
cor(Ed_TermYES,Err_TypeF)        12175
cor(Intercept,Err_TypeL)         11318
cor(CompletenessI,Err_TypeL)     11630
cor(Ed_TermYES,Err_TypeL)        12028
cor(Err_TypeF,Err_TypeL)         11435

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept         3.50      0.67     2.17     4.83 1.00    23490    11258
CompletenessI     0.70      0.46    -0.21     1.59 1.00    28233    11591
Ed_TermYES       -0.35      0.46    -1.26     0.56 1.00    35518    11360
Err_TypeF        -0.39      0.46    -1.30     0.50 1.00    32762    11501
Err_TypeL        -0.02      0.48    -0.96     0.92 1.00    33901    12308

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     9.24      0.66     8.00    10.56 1.00    19455    12334
nu        2.84      0.49     2.05     3.97 1.00    20417    12162

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.10 Summary of the Model Examining Absolute F0 Rise (Hz)

```
Family: student
  Links: mu = identity; sigma = identity; nu = identity
Formula: max_DeltaF0_Hertz ~ 1 + Completeness + Ed_Term + Err_Type + (1 + Completeness
+ Ed_Term + Err_Type | Speaker)
   Data: df_ProsodicMarking (Number of observations: 405)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.24      0.46     0.00     1.50 1.00    19313
sd(CompletenessI)                 1.10      2.52     0.00     9.66 1.00     2618
sd(Ed_TermYES)                    0.32      0.78     0.00     2.39 1.00    20209
sd(Err_TypeF)                     0.25      0.50     0.00     1.60 1.00    22006
```

```
sd(Err_TypeL)                     0.30      0.77     0.00      2.22 1.00    20394
cor(Intercept,CompletenessI)      0.00      0.35    -0.65      0.67 1.00    20585
cor(Intercept,Ed_TermYES)        -0.00      0.36    -0.67      0.66 1.00    28763
cor(CompletenessI,Ed_TermYES)    -0.00      0.36    -0.67      0.68 1.00    20623
cor(Intercept,Err_TypeF)         -0.00      0.36    -0.67      0.67 1.00    30986
cor(CompletenessI,Err_TypeF)     -0.00      0.35    -0.67      0.67 1.00    19930
cor(Ed_TermYES,Err_TypeF)        -0.00      0.35    -0.67      0.67 1.00    14867
cor(Intercept,Err_TypeL)          0.00      0.35    -0.67      0.66 1.00    29757
cor(CompletenessI,Err_TypeL)     -0.00      0.35    -0.66      0.66 1.00    21442
cor(Ed_TermYES,Err_TypeL)         0.00      0.35    -0.67      0.66 1.00    15559
cor(Err_TypeF,Err_TypeL)          0.01      0.35    -0.66      0.68 1.00    13129
                              Tail_ESS
sd(Intercept)                     9215
sd(CompletenessI)                 1795
sd(Ed_TermYES)                    8619
sd(Err_TypeF)                     9741
sd(Err_TypeL)                     8991
cor(Intercept,CompletenessI)     11286
cor(Intercept,Ed_TermYES)        10672
cor(CompletenessI,Ed_TermYES)    10896
cor(Intercept,Err_TypeF)         10629
cor(CompletenessI,Err_TypeF)     11233
cor(Ed_TermYES,Err_TypeF)        10877
cor(Intercept,Err_TypeL)         11439
cor(CompletenessI,Err_TypeL)     11487
cor(Ed_TermYES,Err_TypeL)        11227
cor(Err_TypeF,Err_TypeL)         11613

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept         6.69      1.25     4.13     9.06 1.00     7106     5671
CompletenessI     0.43      0.49    -0.54     1.38 1.00    22629    11122
Ed_TermYES       -0.17      0.49    -1.13     0.78 1.00    30618    11229
Err_TypeF        -0.25      0.49    -1.21     0.71 1.00    26933    11900
Err_TypeL         0.02      0.49    -0.94     0.99 1.00    29530    10757

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    17.99      1.25    15.60    20.51 1.00     9750    10804
nu        2.62      0.41     1.94     3.55 1.00    12648    10929

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.11 Summary of the Model Examining Speech Rate on Segments per second

```
Family: student
  Links: mu = identity; sigma = identity; nu = identity
Formula: Delta_Speech_Rate_Seg ~ 1 + Completeness + Ed_Term + Err_Type + (1 +
Completeness + Ed_Term + Err_Type | Speaker)
   Data: df_filtered_SpeechRate (Number of observations: 400)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                     0.19      0.26     0.00      1.01 1.00    11175
sd(CompletenessI)                 0.16      0.22     0.00      0.81 1.00    17267
sd(Ed_TermYES)                    0.17      0.24     0.00      0.86 1.00    16462
sd(Err_TypeF)                     0.23      0.35     0.00      1.35 1.00    11420
sd(Err_TypeL)                     0.25      0.52     0.00      1.68 1.00    15480
cor(Intercept,CompletenessI)     -0.01      0.35    -0.67      0.65 1.00    24126
cor(Intercept,Ed_TermYES)        -0.00      0.35    -0.67      0.67 1.00    22369
cor(CompletenessI,Ed_TermYES)    -0.00      0.35    -0.66      0.67 1.00    18057
cor(Intercept,Err_TypeF)          0.00      0.36    -0.66      0.68 1.00    23697
```

```
cor(CompletenessI,Err_TypeF)     -0.00      0.35    -0.68     0.66 1.00    18359
cor(Ed_TermYES,Err_TypeF)         0.01      0.35    -0.66     0.67 1.00    14578
cor(Intercept,Err_TypeL)          0.01      0.35    -0.67     0.67 1.00    23018
cor(CompletenessI,Err_TypeL)     -0.01      0.35    -0.68     0.66 1.00    18813
cor(Ed_TermYES,Err_TypeL)         0.00      0.35    -0.66     0.67 1.00    13691
cor(Err_TypeF,Err_TypeL)          0.01      0.35    -0.66     0.67 1.00    12881
                               Tail_ESS
sd(Intercept)                      8659
sd(CompletenessI)                  9315
sd(Ed_TermYES)                     8407
sd(Err_TypeF)                      8760
sd(Err_TypeL)                      8162
cor(Intercept,CompletenessI)      11740
cor(Intercept,Ed_TermYES)         10874
cor(CompletenessI,Ed_TermYES)     11468
cor(Intercept,Err_TypeF)          11310
cor(CompletenessI,Err_TypeF)      11552
cor(Ed_TermYES,Err_TypeF)         12470
cor(Intercept,Err_TypeL)          11148
cor(CompletenessI,Err_TypeL)      10983
cor(Ed_TermYES,Err_TypeL)         11019
cor(Err_TypeF,Err_TypeL)          12063

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept         2.02      0.45     1.15     2.91 1.00    24040    11867
CompletenessI    -0.33      0.40    -1.09     0.46 1.00    26223    10760
Ed_TermYES       -0.04      0.41    -0.84     0.77 1.00    25849    11400
Err_TypeF        -0.11      0.40    -0.88     0.68 1.00    25575    11860
Err_TypeL        -0.03      0.45    -0.91     0.85 1.00    25785    11955

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     5.82      0.30     5.22     6.42 1.00    15846    11520
nu        8.26      2.57     4.62    14.56 1.00    17866    12542

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

E.12 Summary of the Model Examining Speech Rate on Syllables per second

```
Family: student
  Links: mu = identity; sigma = identity; nu = identity
Formula: Delta_Speech_Rate_Syll ~ 1 + Completeness + Ed_Term + Err_Type + (1 +
Completeness + Ed_Term + Err_Type | Speaker)
   Data: df_filtered_SpeechRate (Number of observations: 400)
  Draws: 4 chains, each with iter = 8000; warmup = 4000; thin = 1;
         total post-warmup draws = 16000

Multilevel Hyperparameters:
~Speaker (Number of levels: 36)
                               Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
sd(Intercept)                      0.11      0.11     0.00     0.43 1.00    12837
sd(CompletenessI)                  0.12      0.14     0.00     0.52 1.00    15385
sd(Ed_TermYES)                     0.13      0.14     0.00     0.53 1.00    17691
sd(Err_TypeF)                      0.15      0.17     0.00     0.66 1.00    12545
sd(Err_TypeL)                      0.22      0.32     0.00     1.22 1.00    12311
cor(Intercept,CompletenessI)     -0.02      0.35    -0.68     0.66 1.00    27645
cor(Intercept,Ed_TermYES)        -0.01      0.36    -0.68     0.66 1.00    29044
cor(CompletenessI,Ed_TermYES)    -0.01      0.35    -0.68     0.66 1.00    20835
cor(Intercept,Err_TypeF)         -0.00      0.35    -0.67     0.67 1.00    26801
cor(CompletenessI,Err_TypeF)     -0.00      0.35    -0.67     0.66 1.00    20513
cor(Ed_TermYES,Err_TypeF)        -0.01      0.35    -0.67     0.65 1.00    14347
cor(Intercept,Err_TypeL)          0.01      0.35    -0.67     0.67 1.00    29128
cor(CompletenessI,Err_TypeL)     -0.00      0.35    -0.67     0.66 1.00    20793
cor(Ed_TermYES,Err_TypeL)         0.01      0.35    -0.67     0.66 1.00    14936
cor(Err_TypeF,Err_TypeL)          0.00      0.36    -0.67     0.67 1.00    13191
                               Tail_ESS
```

```
sd(Intercept)                    9124
sd(CompletenessI)                9094
sd(Ed_TermYES)                   9667
sd(Err_TypeF)                    9215
sd(Err_TypeL)                    9181
cor(Intercept,CompletenessI)    11303
cor(Intercept,Ed_TermYES)       12163
cor(CompletenessI,Ed_TermYES)   11938
cor(Intercept,Err_TypeF)        10508
cor(CompletenessI,Err_TypeF)    12241
cor(Ed_TermYES,Err_TypeF)       12445
cor(Intercept,Err_TypeL)        10508
cor(CompletenessI,Err_TypeL)    11484
cor(Ed_TermYES,Err_TypeL)       12250
cor(Err_TypeF,Err_TypeL)        12581

Regression Coefficients:
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept         0.93      0.26     0.43     1.45 1.00    30986    11404
CompletenessI    -0.29      0.27    -0.81     0.23 1.00    32138    11038
Ed_TermYES        0.00      0.28    -0.55     0.56 1.00    29725    11697
Err_TypeF        -0.16      0.27    -0.69     0.37 1.00    27527    12549
Err_TypeL         0.02      0.35    -0.68     0.71 1.00    27976    12111

Further Distributional Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     2.71      0.15     2.42     3.01 1.00    17577    12113
nu        7.06      2.09     4.07    12.12 1.00    18887    13294

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```